



The Erasmus Computing Grid



Building a Super-Computer
for *Free*

Dr. Tobias A. Knoch & Luc V. de Zeeuw
with Bert Eussen & Michael Moorhouse

Cell Biology & Genetics, Clinical Genetics, Virology, Erasmus MC

Hogeschool Rotterdam, Rotterdam, The Netherlands

ECG Basic Grid Structure

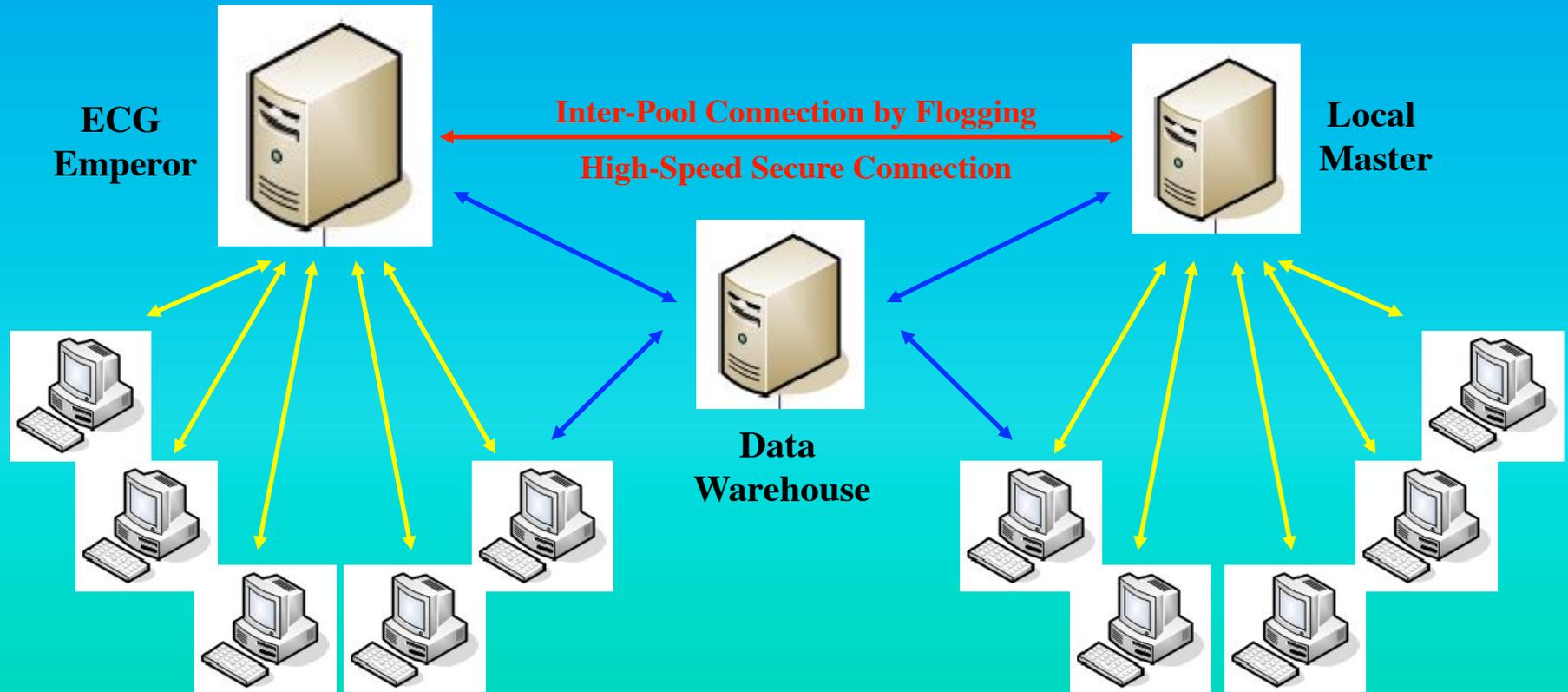


The ECG consists of the computer pools of the Erasmus MC and the Hogeschool Rotterdam. The client computers are controlled by the general ECG Emperor of the grid at EMC and a local master at HR. "Condor" is used as middleware since it is very well established and open-source.



~9000 CPUs or ~18000 vHosts

~3500 CPUs or ~7000 vHosts



ECG Basic Grid Structure

Erasmus Computing Grid



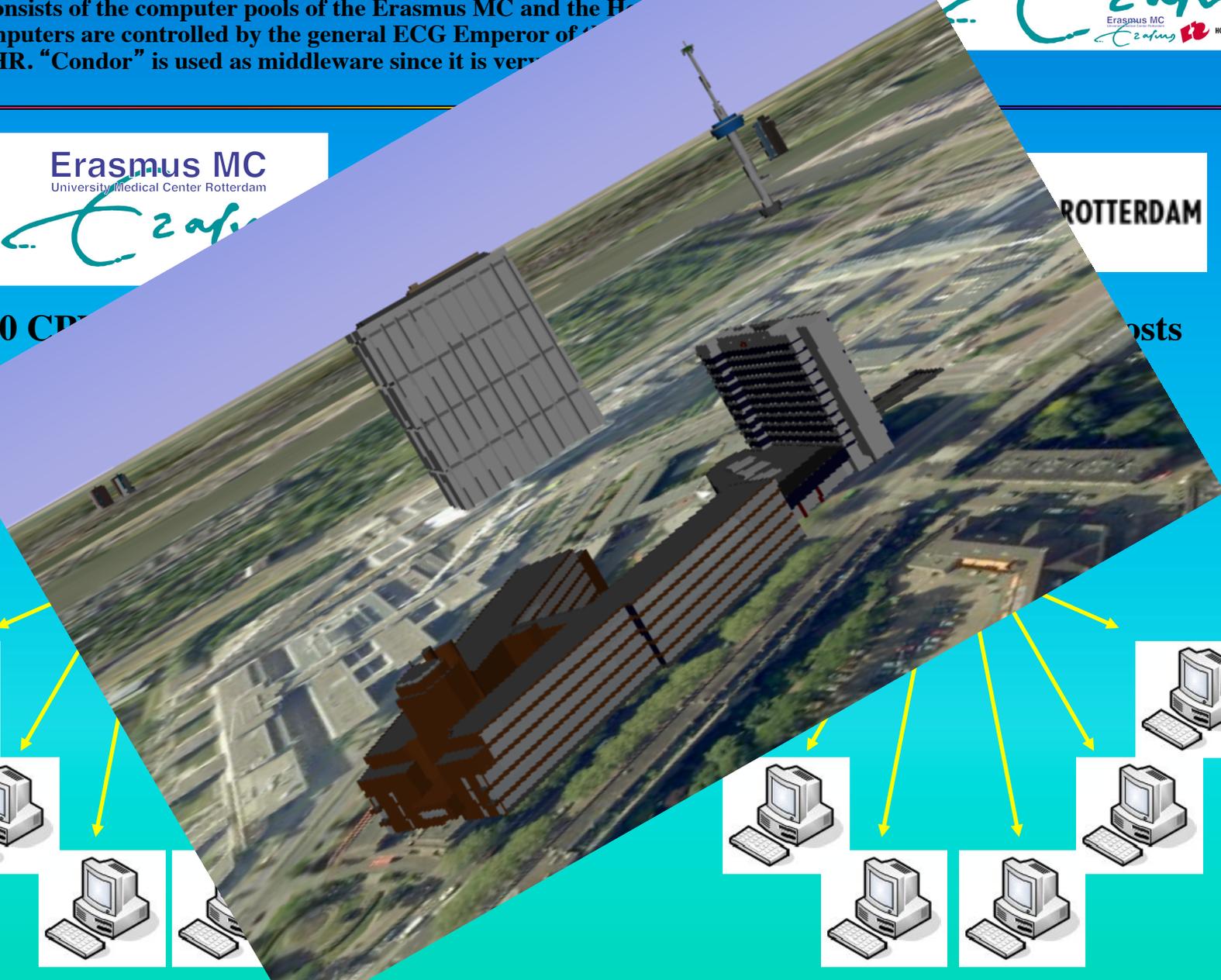
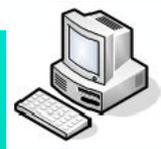
The ECG consists of the computer pools of the Erasmus MC and the HR. The client computers are controlled by the general ECG Emperor of the HR. "Condor" is used as middleware since it is very



~9000 CPUs

costs

ECG Emperor

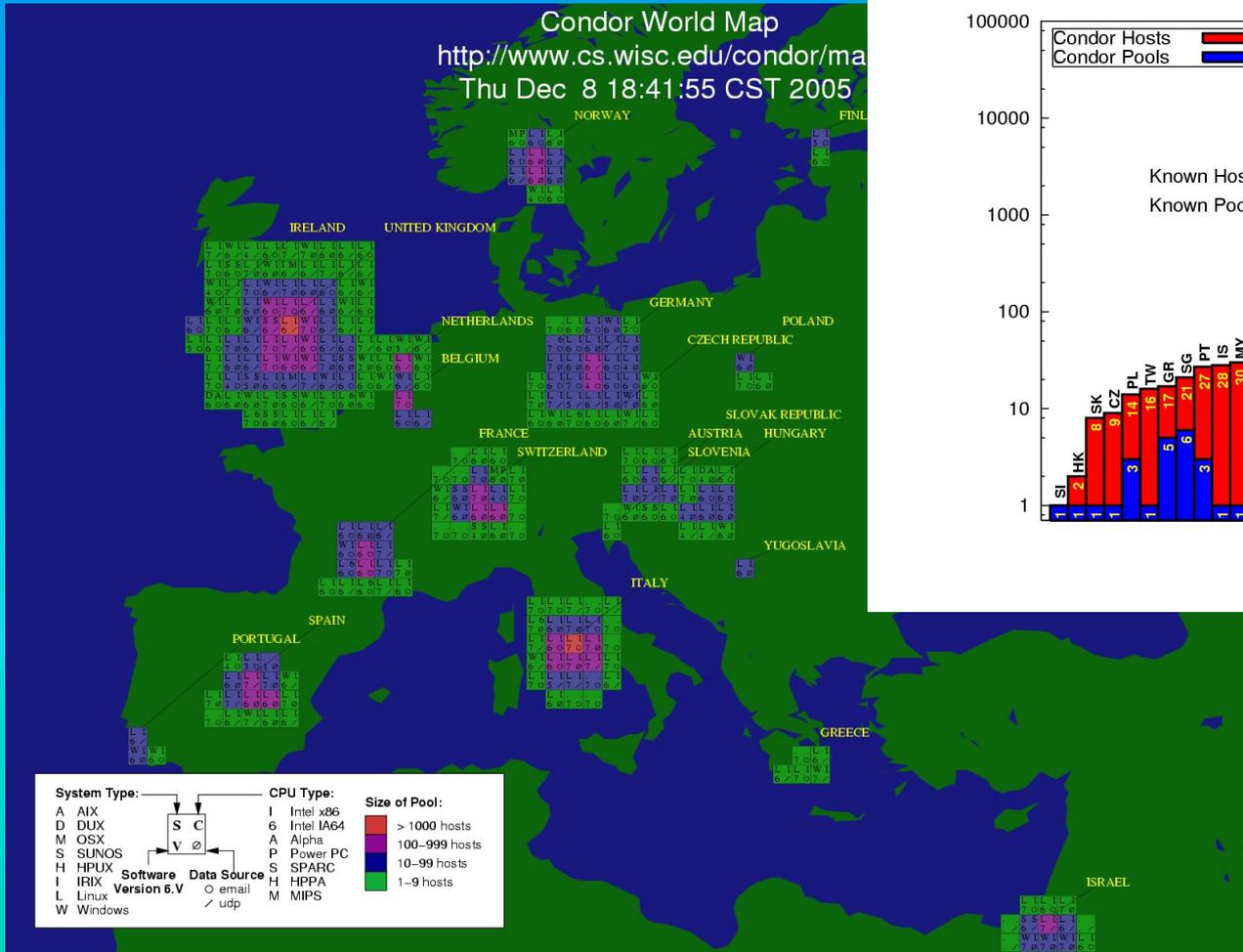


World-Map of Condor Grids

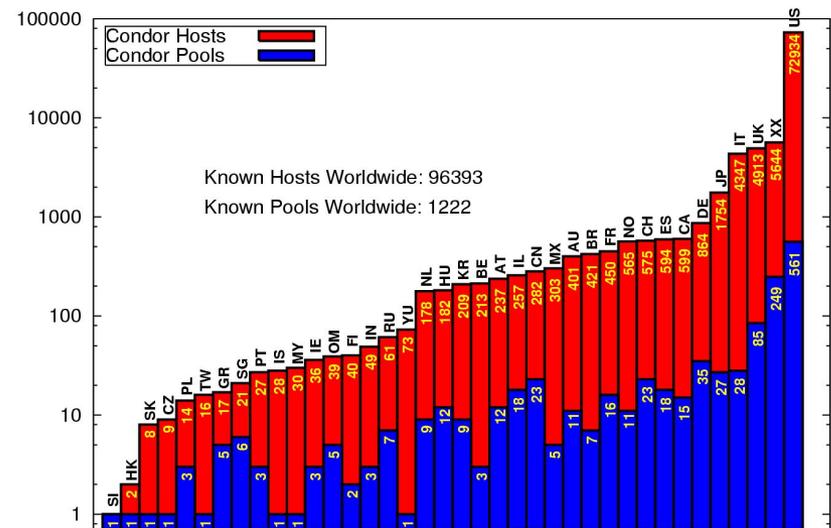


Condor is used as the standard for batch systems on many real parallel super-computers as well as in many distributed trivial parallel computing grids. It is running in very diverse environments of the highest security levels as e.g. governmental or industry R&D.

With a potential capacity of currently ~20 to 25 Tera flops the ECG is one of the biggest grids!



Known Condor Pools and Hosts by Country
 Sun Dec 11 18:40:44 CST 2005



Based on pool operator reports only!

Installation and Security



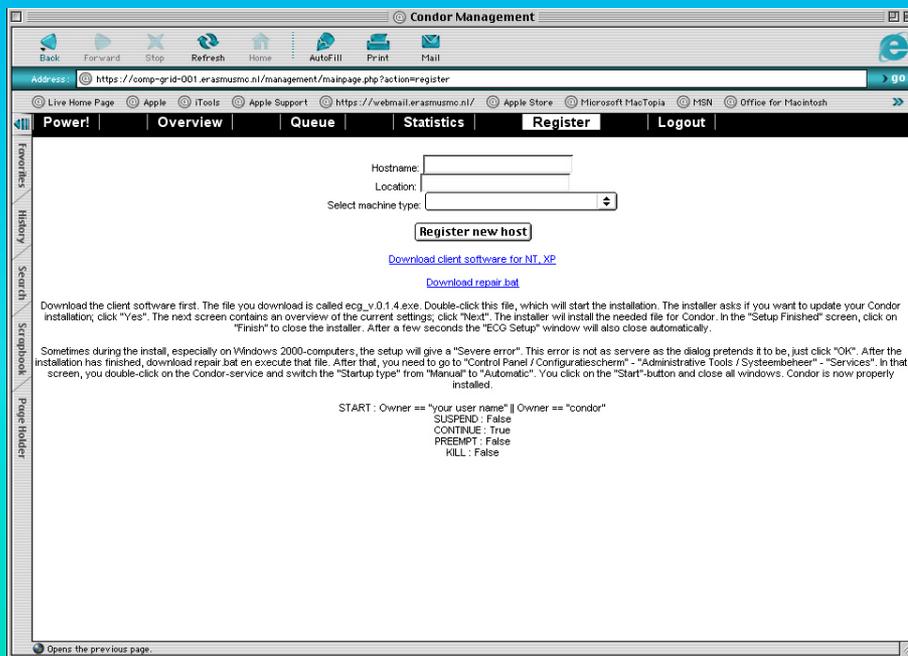
For installation, Condor is installed administratively via a secure login from the ECG management site using the corresponding installers.

Condor is an established middleware with no known risks: Condor is running as a guest, no data can be accessed in any form, and nothing can be changed on the local client computer!

Installation

2 Minutes per computer using the ECG management system

- ❖ software install
- ❖ registration
- ❖ general maintenance



Security

The ECG is NO Risk to the client computer or EMC/HR

- ❖ runs as a *guest*
- ❖ data connection to master is *secure*
- ❖ no local data are visible to condor
- ❖ the local system *cannot* be changed
- ❖ application executables are *certified*
- ❖ input data is *certified* by ECG
- ❖ local user *always* has usage priority
- ❖ runs *only* if vhosts are available
- ❖ jobs can be stopped *always* automatically
- ❖ ECG can be used as a security monitoring system (e.g. power outage or flooding)
- ❖ **PRIVACY RULES ARE APPLIED AND GUARANTIED!!!**

ECG - Management

To maintain the ECG a data base with installed and registered hosts was created. It allows to track the status of the local clients and gives global statistical data about performance. A job-submission and accounting system is currently implemented. The ECG is also working together with the EGEE and e-Science grids and is a partner of the German nationwide MediGRID / D-Grid.



Condor Management

Address: <https://comp-grid-001.erasmusmc.nl/management/mainpage.php?action=status&display=table>

Live Home Page Apple iTools Apple Support https://webmail.erasmusmc.nl/ Apple Store Microsoft MacTopia MSN Office for Macintosh

vm1@TigerLion	10.98.4.7	695.16 MFlops	511 MB	0	2005-11-07 13:20:19.777098	[x][i]
vm2@TigerLion	10.98.4.7	695.16 MFlops	511 MB	0	2005-11-07 13:20:19.977908	[x][i]
vm3@TigerLion					11-07 13:20:20.186658	[x][i]
vm4@TigerLion					11-07 13:20:20.389709	[x][i]
vm1@bioinf-mir.erasmusmc.nl					11-07 13:20:01.606706	[x][i]
vm2@bioinf-mir.erasmusmc.nl					11-07 13:20:01.801927	[x][i]
vm3@bioinf-mir.erasmusmc.nl					11-07 13:20:01.996655	[x][i]
vm4@bioinf-mir.erasmusmc.nl					11-07 13:20:02.194894	[x][i]
vm1@cbg-grid.gen.fgg.eur.nl					11-07 13:20:02.404949	[x][i]
vm2@cbg-grid.gen.fgg.eur.nl					11-07 13:20:02.60134	[x][i]
vm1@cbg3.erasmusmc.nl					11-07 13:20:02.887505	[x][i]
vm2@cbg3.erasmusmc.nl					11-07 13:20:03.086038	[x][i]
vm3@cbg3.erasmusmc.nl					11-07 13:20:03.282584	[x][i]
vm4@cbg3.erasmusmc.nl					11-07 13:20:03.481282	[x][i]
ch1-42402679					11-07 13:20:05.17493	[x][i]
dell1032221204					11-07 13:20:05.369733	[x][i]
dell103233481	10.98.25.63	712.76 MFlops	509 MB	0	2005-11-07 13:20:05.566907	[x][i]

A circular gauge with a scale from 0 to 120. A red star is positioned at the 50 mark, with a red arrow pointing to it and the text "~5000 vhosts". A red box at the bottom of the gauge contains the text "~3000.0 GigaFlops/cluster". A "Login" button is located at the top of the gauge.

ECG - Business Concepts

The ECG has four major business concepts: the research, education, diagnostics and industry arenas. For all of these, separate business concepts, work-flows and user managements are necessary and need to be integrated to allow maximum efficiency.

The ECG is one of the few grids on this level of complexity !



Research:

- ❖ genomic and proteomic analysis
- ❖ epidemiology
- ❖ image analysis, e.g. Applied Molecular Imaging (AMI)

Education:

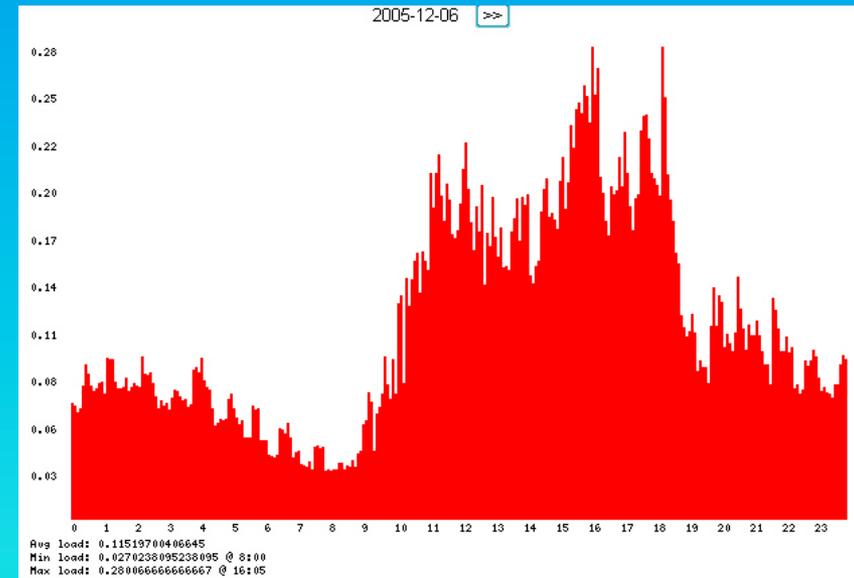
- ❖ training of the coming grid generation of IT specialists
- ❖ developing new concepts for grid computing

Diagnostics:

- ❖ clinical image and data analysis
- ❖ operation planning and operation support

Industry:

- ❖ brokerage of computing resources



Acknowledgements



Workforce:

Rob de Graaf

Remko Kleinjan

Redmar van Leeuwarden

Sander van Vliet

Anis Abuseiris, GLOBE-Consortium ErasmusMC

Dr. Michael J. Moorhouse, GLOBE-Consortium ErasmusMC

Dr. Hubert J. Eussen, GLOBE-Consortium ErasmusMC

Steering Committee:

Prof. Dr. Frank G. Grosveld, ErasmusMC

Prof. Dr. Hans van den Berg, HR

Braam van Laar, ErasmusMC

Rob Ketelaar, HR

Institutions:

Erasmus Medical Center

Hogeschool Rotterdam

The German MediGRID and D-Grid Initiatives

The German Ministry for Science and Technology

Acknowledgement



Workforce:

Rob de Graaf

Remko Kleinjan

Redmar

Sam

Pr

Prof

Braam

Rob Ket

Institution

Erasmus Medic

Hogeschool Rott

The German Medic

The German Ministry

Future Ambition:

Griddify the Rotterdam Region &

The Netherlands

The Erasmus Computing Grid

-

Building a Super-Computer Virtually for Free

at the

Erasmus Medical Center and the Hogeschool Rotterdam

Knoch, T. A., de Zeeuw, L. J.

Desktop Grid Experience workshop, AlmereGrid Initiative, Almere, The Netherlands, 28th September, 2006.

Abstract

The Set-Up of the 20 Teraflop Erasmus Computing Grid: To meet the enormous computational needs of life-science research as well as clinical diagnostics and treatment the Hogeschool Rotterdam and the Erasmus Medical Center are currently setting up one of the largest desktop computing grids in the world – The Erasmus Computing Grid. Currently 3 Tera flops are operational and in early production, installation up to the today available maximum capacity of 20 Tera flops in both institutions is planned and partly underway. Thus the Erasmus Computing Grid transforms the existing and sustained huge computer capacity available into usable form via a reliable and secure installing and management system, so that the academic and industrial opportunities depending on such huge computing capacities can be realized for the benefit of society.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:

Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, coevolution, holistic genetics, genome mechanics, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension,

local nuclear dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization, confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, auto-fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual paper tool.

Literature References

- Knoch, T. A.** Dreidimensionale Organisation von Chromosomen-Domänen in Simulation und Experiment. (Three-dimensional organization of chromosome domains in simulation and experiment.) *Diploma Thesis*, Faculty for Physics and Astronomy, Ruperto-Carola University, Heidelberg, Germany, 1998, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-010685-5 and ISBN 978-3-00-010685-9 (soft cover, 2nd ed.), ISBN 3-00-035857-9 and ISBN 978-3-00-0358857-0 (hard cover, 2nd ed.), ISBN 3-00-035858-7, and ISBN 978-3-00-035858-6 (DVD, 2nd ed.), 1998.
- Knoch, T. A.,** Münkkel, C. & Langowski, J. Three-dimensional organization of chromosome territories and the human cell nucleus - about the structure of a self replicating nano fabrication site. *Foresight Institute - Article Archive*, Foresight Institute, Palo Alto, CA, USA, <http://www.foresight.org>, 1- 6, 1998.
- Knoch, T. A.,** Münkkel, C. & Langowski, J. Three-Dimensional Organization of Chromosome Territories and the Human Interphase Nucleus. *High Performance Scientific Supercomputing*, editor Wilfried Juling, Scientific Supercomputing Center (SSC) Karlsruhe, University of Karlsruhe (TH), 27- 29, 1999.
- Knoch, T. A.,** Münkkel, C. & Langowski, J. Three-dimensional organization of chromosome territories in the human interphase nucleus. *High Performance Computing in Science and Engineering 1999*, editors Krause, E. & Jäger, W., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.
- Bestvater, F., **Knoch, T. A.,** Langowski, J. & Spiess, E. GFP-Walking: Artificial construct conversions caused by simultaneous cotransfection. *BioTechniques* 32(4), 844-854, 2002.
- Knoch, T. A. (editor),** Backes, M., Baumgärtner, V., Eysel, G., Fehrenbach, H., Göker, M., Hampl, J., Hampl, U., Hartmann, D., Hitzelberger, H., Nambena, J., Rehberg, U., Schmidt, S., Weber, A., & Weidemann, T. Humanökologische Perspektiven Wechsel - Festschrift zu Ehren des 70. Geburtstags von Prof. Dr. Kurt Egger. Human Ecology Working Group, Ruperto-Carola University of Heidelberg, Heidelberg, Germany, 2002.
- Knoch, T. A.** Approaching the three-dimensional organization of the human genome: structural-, scaling- and dynamic properties in the simulation of interphase chromosomes and cell nuclei, long- range correlations in complete genomes, *in vivo* quantification of the chromatin distribution, construct conversions in simultaneous co-transfections. *Dissertation*, Ruperto-Carola University, Heidelberg, Germany, and TAK†Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-009959-X and ISBN 978-3-00-009959-5 (soft cover, 3rd ed.), ISBN 3-00-009960-3 and ISBN 978-3-00-009960-1 (hard cover, 3rd ed.), ISBN 3-00-035856-9 and ISBN 978-3-00-010685-9 (DVD, 3rd ed.) 2002.
- Knoch, T. A.** Towards a holistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. *High Performance Computing in Science and Engineering 2003*, editors Krause, E., Jäger, W. & Resch, M., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3- 540-40850-9, 421-440, 2003.
- Wachsmuth, M., Weidemann, T., Müller, G., Urs W. Hoffmann-Rohrer, **Knoch, T. A.,** Waldeck, W. & Langowski, J. Analyzing intracellular binding and diffusion with continuous fluorescence photobleaching. *Biophys. J.* 84(5), 3353-3363, 2003.

- Weidemann, T., Wachsmuth, M., **Knoch, T. A.**, Müller, G., Waldeck, W. & Langowski, J. Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *J. Mol. Biol.* 334(2), 229-240, 2003.
- Fejes Tóth, K., **Knoch, T. A.**, Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C. P., Müller, G. & Rippe, K. Trichostatin A induced histone acetylation causes decondensation of interphase chromatin. *J. Cell Science* 177, 4277-4287, 2004.
- Ermler, S., Kronic, D., **Knoch, T. A.**, Moshir, S., Mai, S., Greulich-Bode, K. M. & Boukamp, P. Cell cycle-dependent 3D distribution of telomeres and telomere repeat-binding factor 2 (TRF2) in HaCaT and HaCaT-myc cells. *Europ. J. Cell Biol.* 83(11-12), 681-690, 2004.
- Kost, C., Gama de Oliveira, E., **Knoch, T. A.** & Wirth, R. Spatio-temporal permanence and plasticity of foraging trails in young and mature leaf-cutting ant colonies (*Atta spp.*). *J. Trop. Ecol.* 21(6), 677- 688, 2005.
- Winnefeld, M., Grewenig, A., Schnölzer, M., Spring, H., **Knoch, T. A.**, Gan, E. C., Rommelaere, J. & Cziepluch, C. Human SGT interacts with BAG-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Exp. Cell Res.* 312, 2500-2514, 2006.