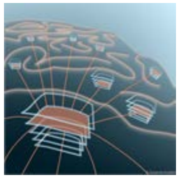


Biological neuronal network simulation code for the brain scale

January 11, 2012 | **Markus Diesmann**

Institute of Neuroscience and Medicine (INM-6)
Computational and Systems Neuroscience

Institute for Advanced Simulation (IAS-6)
Theoretical Neuroscience

Blue Gene Active Storage Workshop

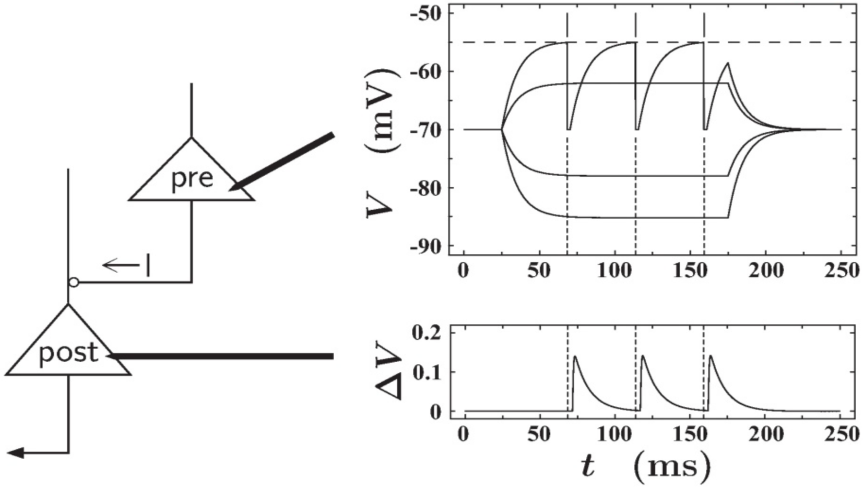
Outline

- fundamental neuronal interactions
- example of model construction: local cortical network
- critique of local network models
- necessity and challenges of brain-scale models
- scalability of neuronal network simulations
- state-of-the-art and comparison of K and JUGENE
- i/o challenges for simulations at the brain-scale

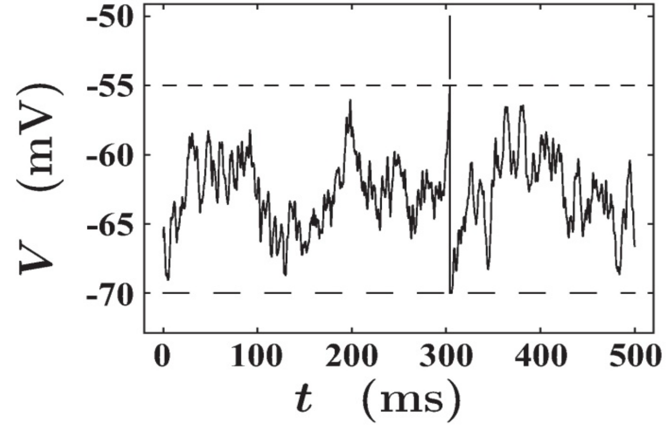
nest::initiative JÜLICH FORSCHUNGSZENTRUM

Fundamental interactions

in vitro



in vivo



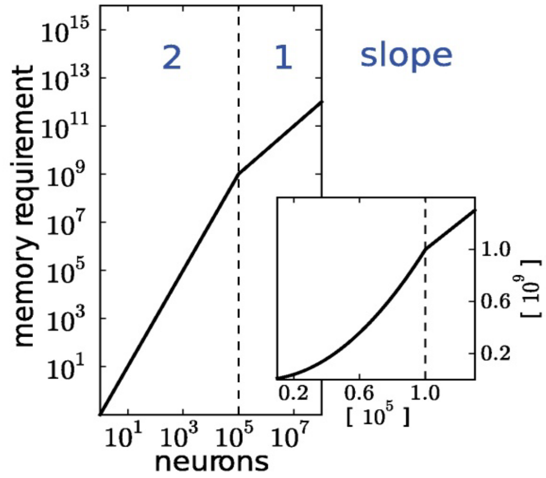
- current injection into pre-synaptic neuron causes excursions of membrane potential
- supra-threshold value causes spike transmitted to post-synaptic neuron
- post-synaptic neuron responds with small excursion of potential after delay
- inhibitory neurons (20%) cause negative excursion

- each neuron receives input from 10,000 other neurons
- causing large fluctuations of membrane potential
- emission rate of 1 to 10 spikes per second

Fundamental interactions

- connectivity $c = 0.1$
- synapses per neuron $= 10^4$

$\Rightarrow$ minimal network size $= 10^5$



network $N = 10^5$

- considered elementary unit
- corresponding to 1 mm^3

$\Rightarrow$ possible

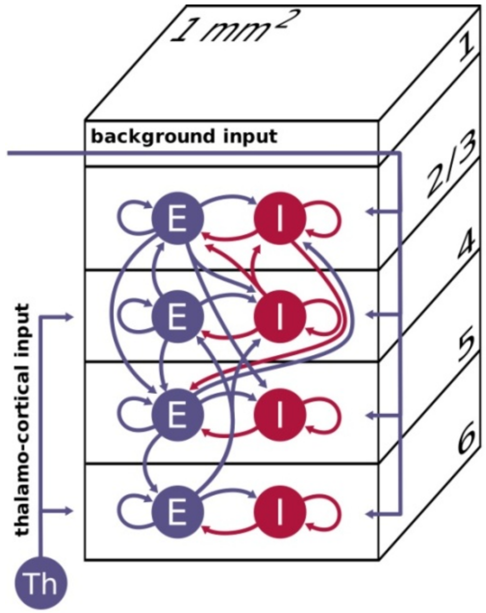
total number of synapses $= (cN) \cdot N$

Morrison, Mehring, Geisel, Aertsen, Diesmann (2005) *Neural Computation* 17:1776–1801

Morrison, Straube, Plesser, Diesmann (2007) *Neural Computation* 19:47–79

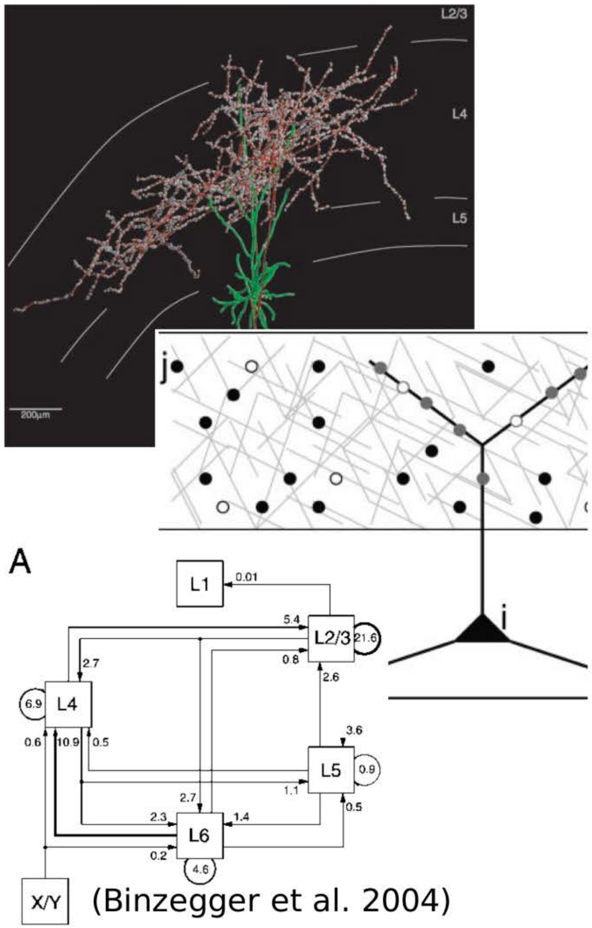
Minimal layered cortical network model

- 1 mm^3
- 1 billion synapses, 100,000 neurons
- 2 populations of neurons (E,I) per layer
- E and I identical neuronal dynamics
- laterally homogeneous connectivity
- layer- and type-specific C_{ij}^{xy}



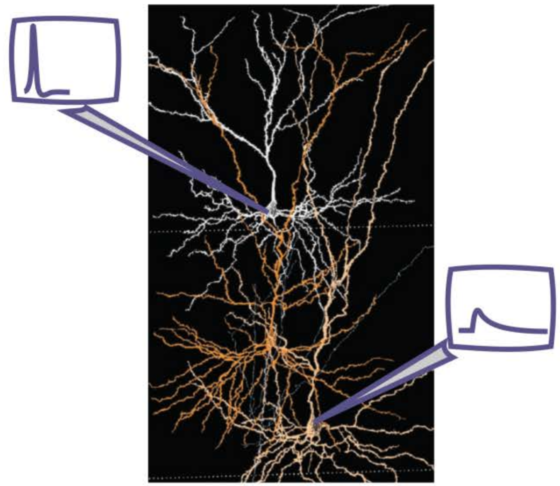
Available data sets

in vivo anatomy



(Binzegger et al. 2004)

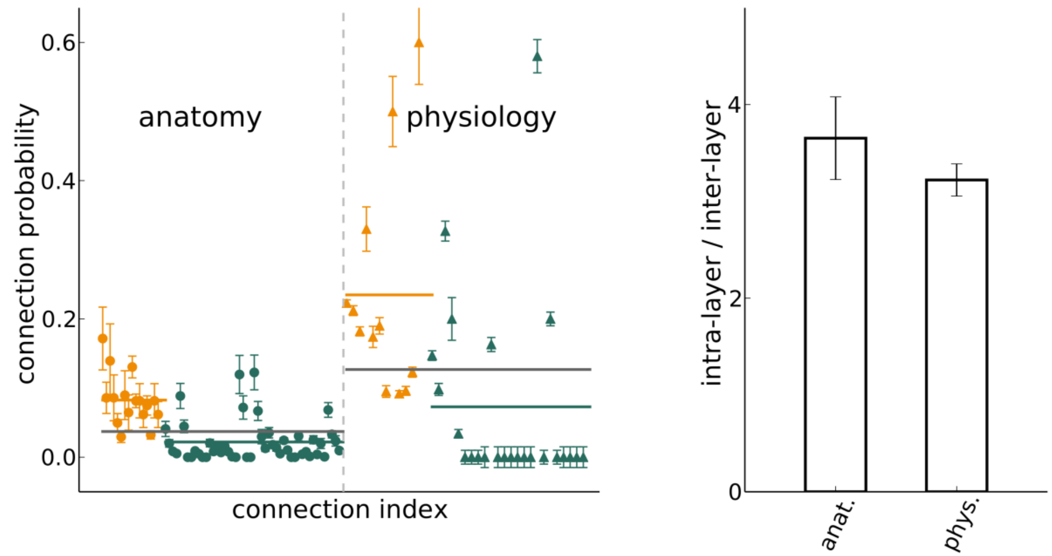
in vitro physiology



Type of connection[a]	Connectivity ratio
L5 pyramid to L5 pyramid	1:11 (15:163)
L2/3 pyramid to L2/3 pyramid	1:4 (65:247)
	1:10 (8:81)
L4 excitatory to L4 excitatory	**1:5.7 (4:23)**
L3 pyramid to L5 pyramid	1:1.8 (16:29)
[Postsynaptic apical dendrite]	
	1:1 (2:2)
L5 pyramid to L3 pyramid	1:29
L4 excitatory to L3 pyramid	1:3.6 (7:25)
(Presynaptic spiny stellates) ($n = 4$)	**1:10 (7:70)**
L5 pyramid to L5 interneuron	1:10.4 (7:73)
L5 interneuron to L5 pyramid	*1:8 (9:73)*

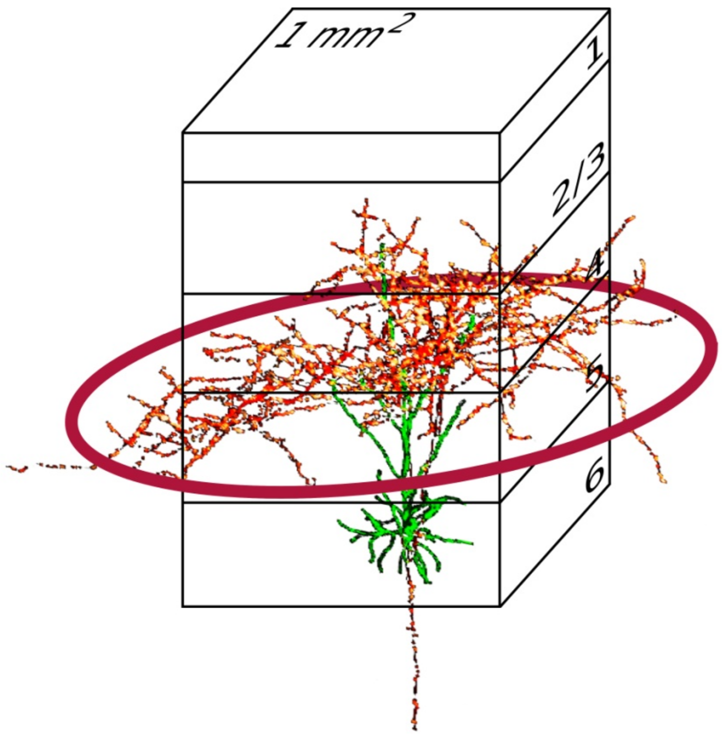
(Thomson et al. 2002)

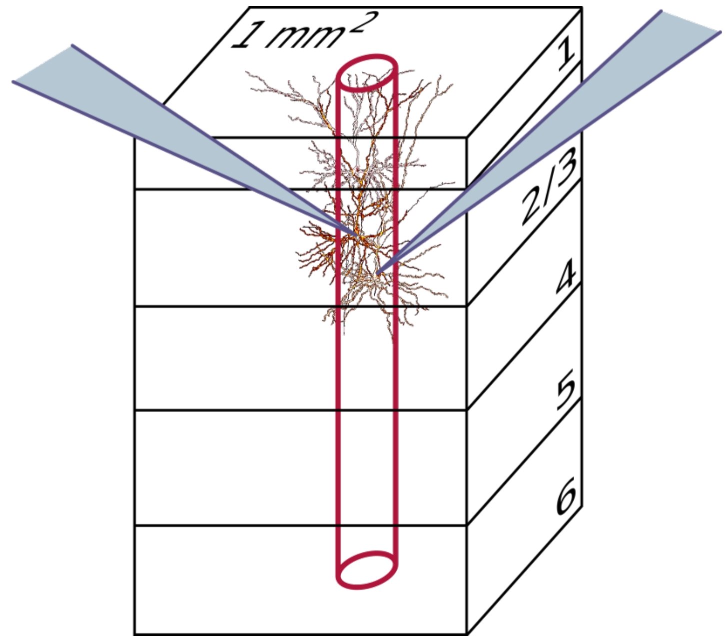
Consistency of connection probabilities



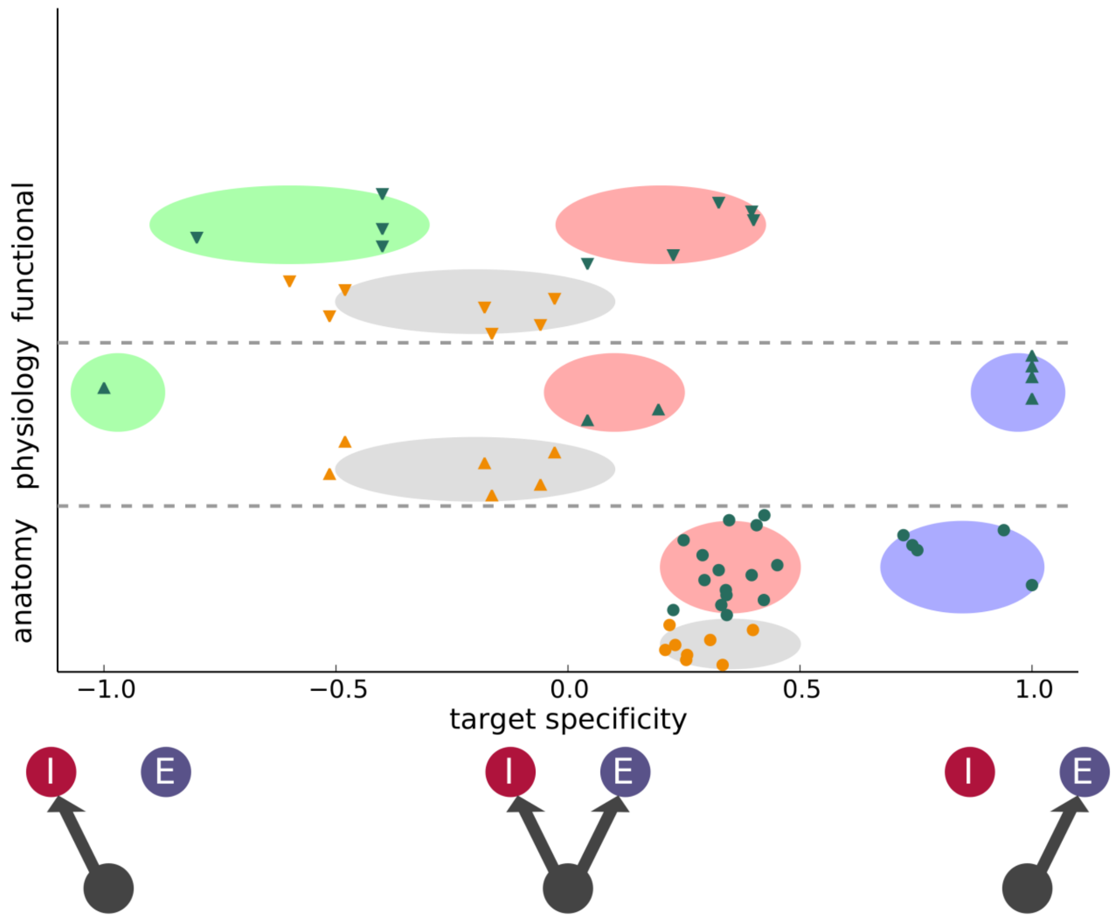
- inconsistent averages
- consistent architectural relations

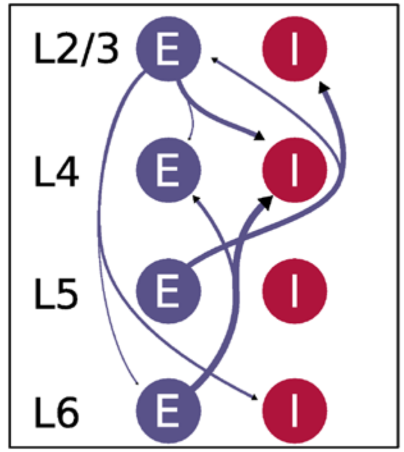
Lateral connectivity



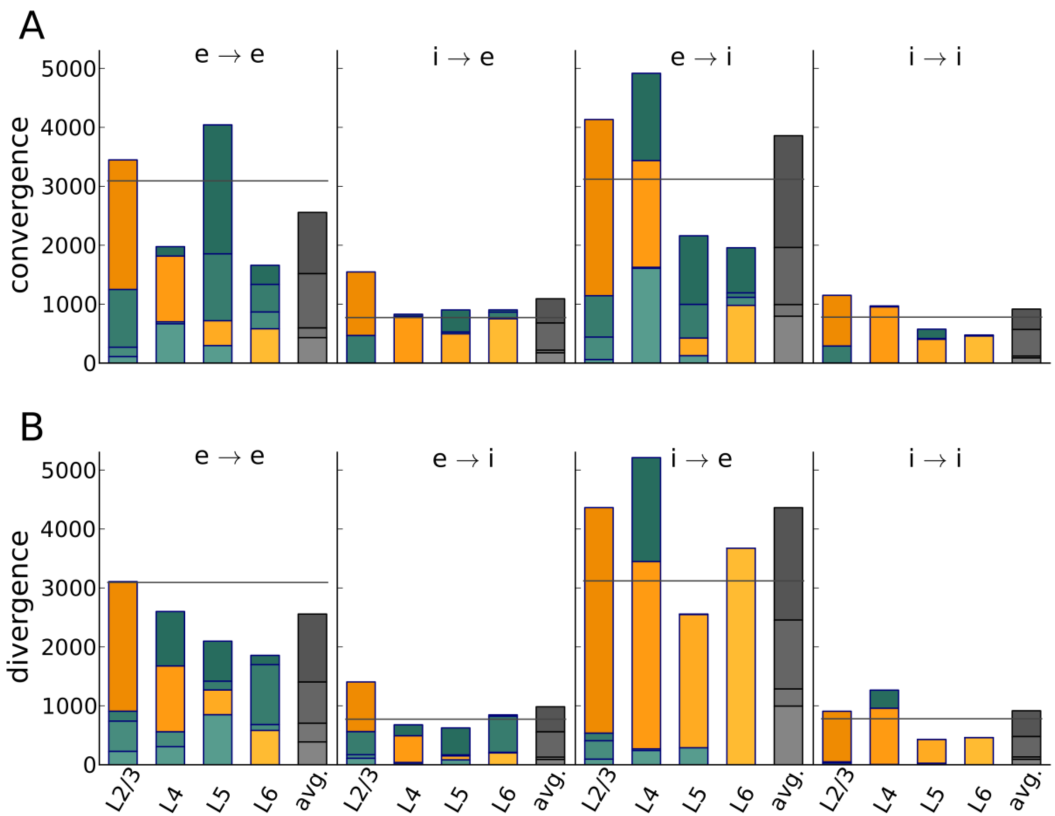


Target specificity





Convergence and divergence



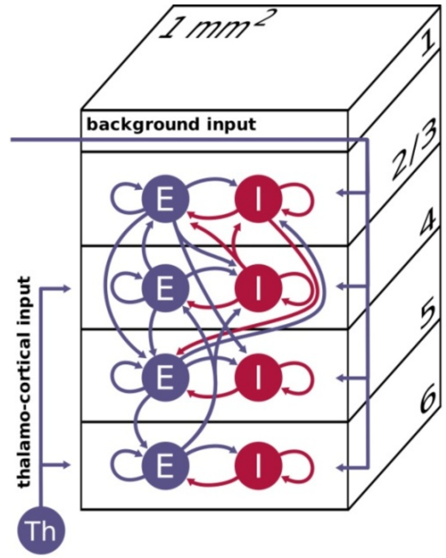
- dominated by within-layer connections
- e → e divergence reflects ”standard” loop
- e → i divergence reflects target-specific feedback

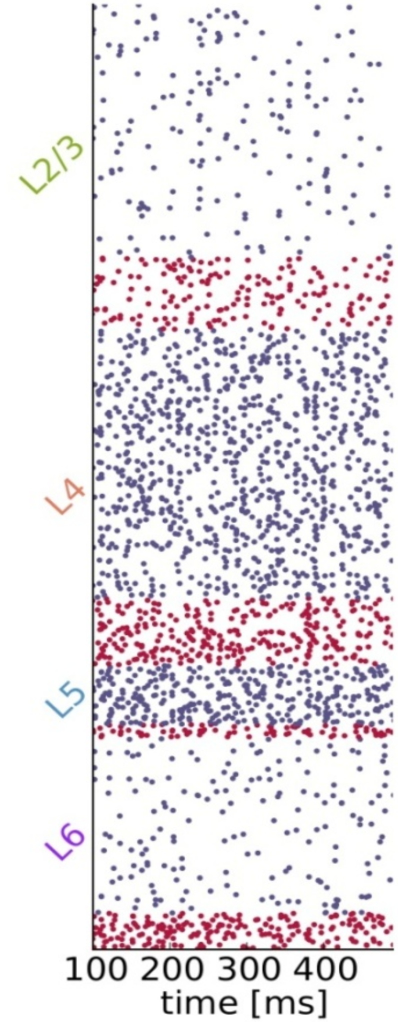
Local cortical microcircuit

- 10^5 neurons and 10^9 synapses
- excitatory (80%) and inhibitory population (20%) of neurons in each layer

taking into account layer and neuron-type specific connectivity is sufficient to reproduce experimentally observed:

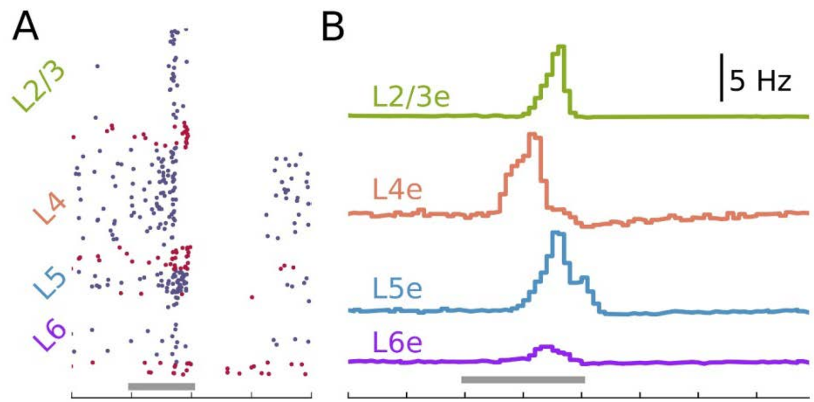
- asynchronous-irregular spiking of neurons
- correct distribution of spike rates across layers

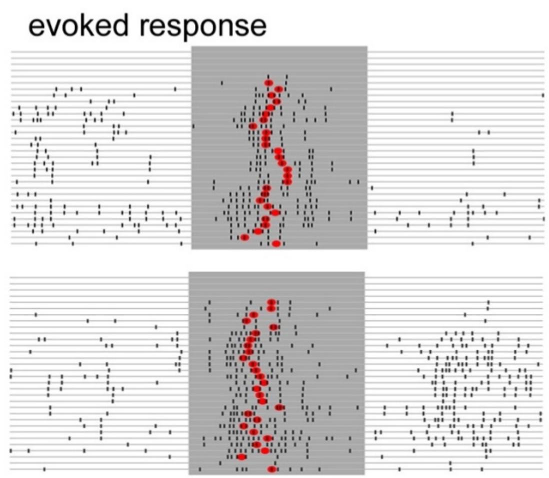


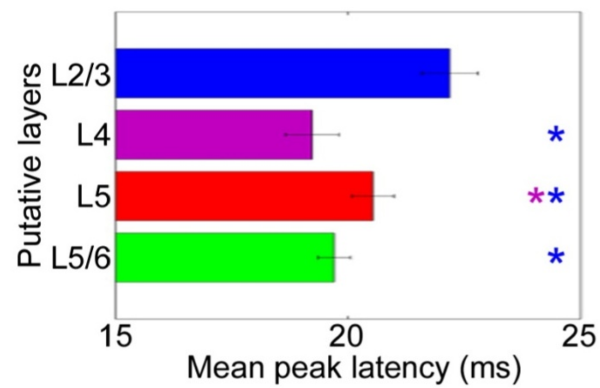


Potjans TC & Diesmann M (2012) The cell-type specific connectivity of the local cortical network explains prominent features of neuronal activity. *Cerebral Cortex* 10.1093/cercor/bhs358

Response to transient inputs





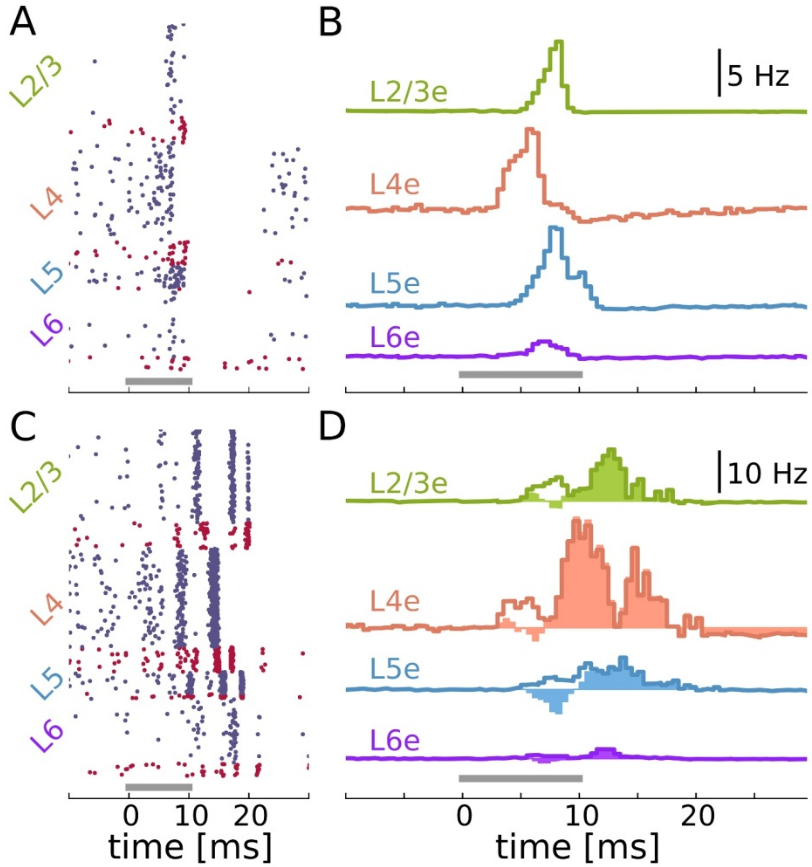


Sakata and Harris (2009) Neuron

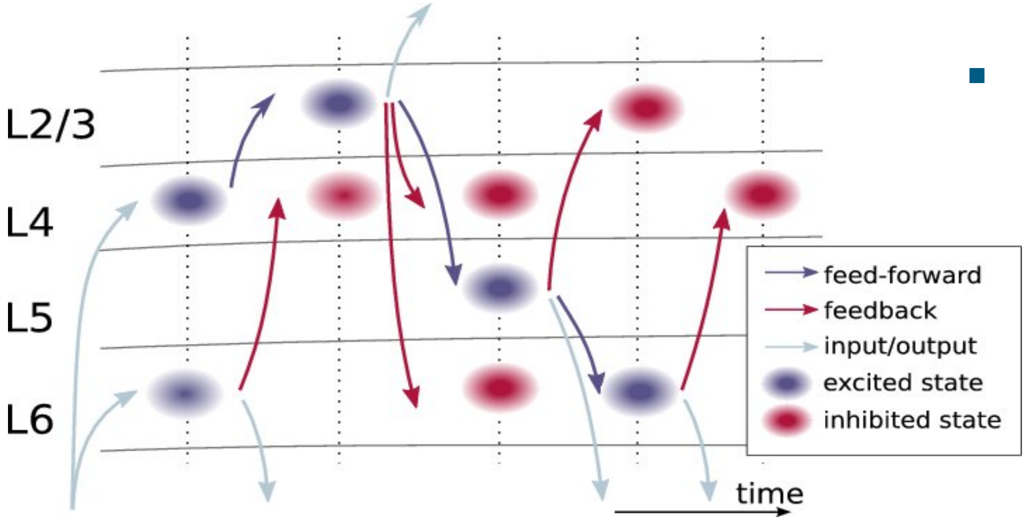
Response to transient inputs

- T = -0.4

- T = +0.4



Hypothesis on cortical flow of activity



- **building block for further studies:** Wagatsuma N, Potjans TC, Diesmann M and Fukai T (2011) Layer-dependent attentional processing by top-down signals in a visual cortical microcircuit. *Front Comput Neurosci* 5:31

The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model

Tobias C. Potjans[1,2,3] and **Markus Diesmann**[1,2,4,5]

Author Affiliations

[1]Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Research Center Juelich, Juelich, Germany
[2]Brain and Neural Systems Team, RIKEN Computational Science Research Program, Wako-shi, Japan
[3]Faculty of Biology III, Albert-Ludwigs-University Freiburg, Freiburg, Germany
[4]Medical Faculty, RWTH Aachen University, Aachen, Germany
[5]RIKEN Brain Science Institute, Wako-shi, Japan

This Article

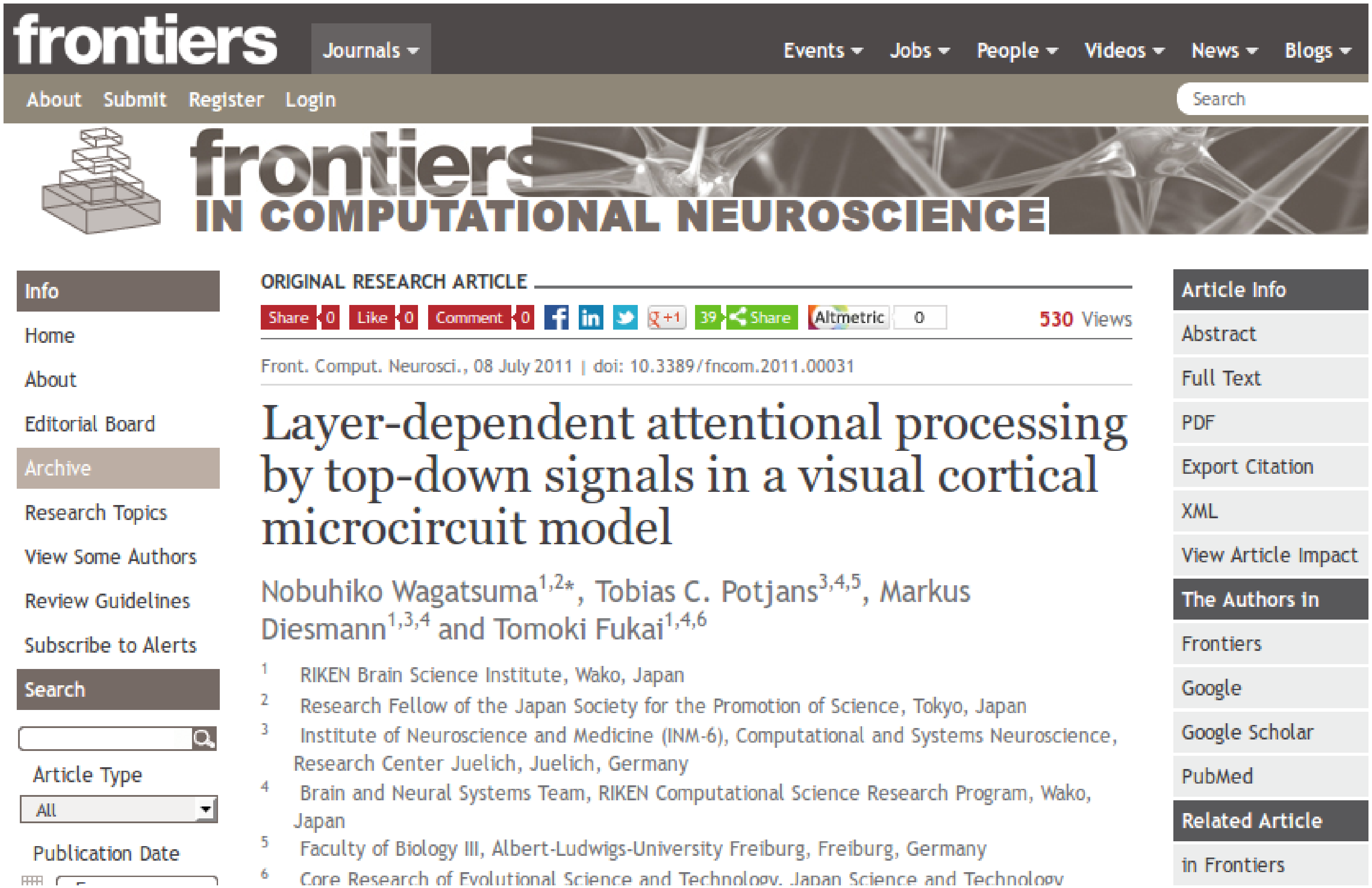
Cereb. Cortex (2012)
doi: 10.1093/cercor/bhs358
First published online: December 2, 2012

This article is Open Access
» Abstract *Free*
Full Text (HTML) *Free*
Full Text (PDF) *Free*

Classifications
Article

Services
Alert me when cited

Cortical flow of activity



frontiers Journals Events Jobs People Videos News Blogs

About Submit Register Login Search

frontiers IN COMPUTATIONAL NEUROSCIENCE

Info
- Home
- About
- Editorial Board
- Archive
- Research Topics
- View Some Authors
- Review Guidelines
- Subscribe to Alerts

Search

Article Type: All

Publication Date

ORIGINAL RESEARCH ARTICLE

Share 0 Like 0 Comment 0 39 Share Altmetric 0 530 Views

Front. Comput. Neurosci., 08 July 2011 | doi: 10.3389/fncom.2011.00031

Layer-dependent attentional processing by top-down signals in a visual cortical microcircuit model

Nobuhiko Wagatsuma[1,2*], Tobias C. Potjans[3,4,5], Markus Diesmann[1,3,4] and Tomoki Fukai[1,4,6]

1 RIKEN Brain Science Institute, Wako, Japan
2 Research Fellow of the Japan Society for the Promotion of Science, Tokyo, Japan
3 Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Research Center Juelich, Juelich, Germany
4 Brain and Neural Systems Team, RIKEN Computational Science Research Program, Wako, Japan
5 Faculty of Biology III, Albert-Ludwigs-University Freiburg, Freiburg, Germany
6 Core Research of Evolutional Science and Technology, Japan Science and Technology

Article Info
- Abstract
- Full Text
- PDF
- Export Citation
- XML
- View Article Impact

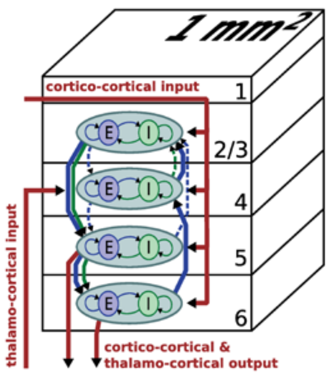
The Authors in
- Frontiers
- Google
- Google Scholar
- PubMed

Related Article
- in Frontiers

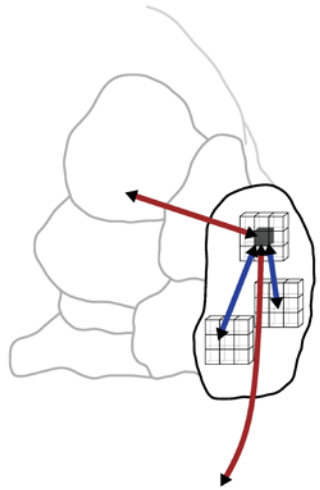
Brain-scale connectivity

[redacted]

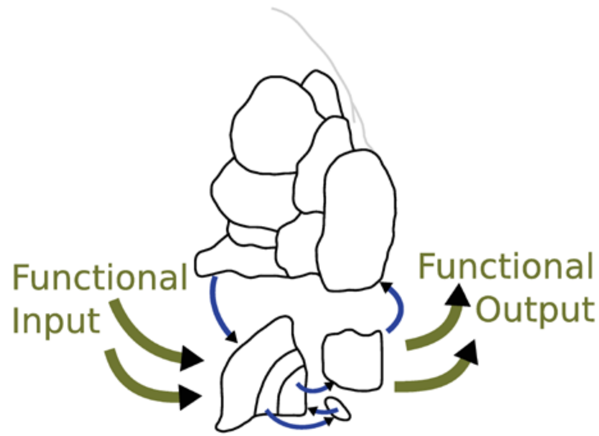
Local Cortical Network Connections:
- ✓ local connections
- ✓ realistic synaptic modeling
- ✗ a major part of synapses missing
- ⇒ input dependent local network dynamics



Cortical Area Network Connections:
- ✓ intrinsic connections
- ✗ many synapses missing



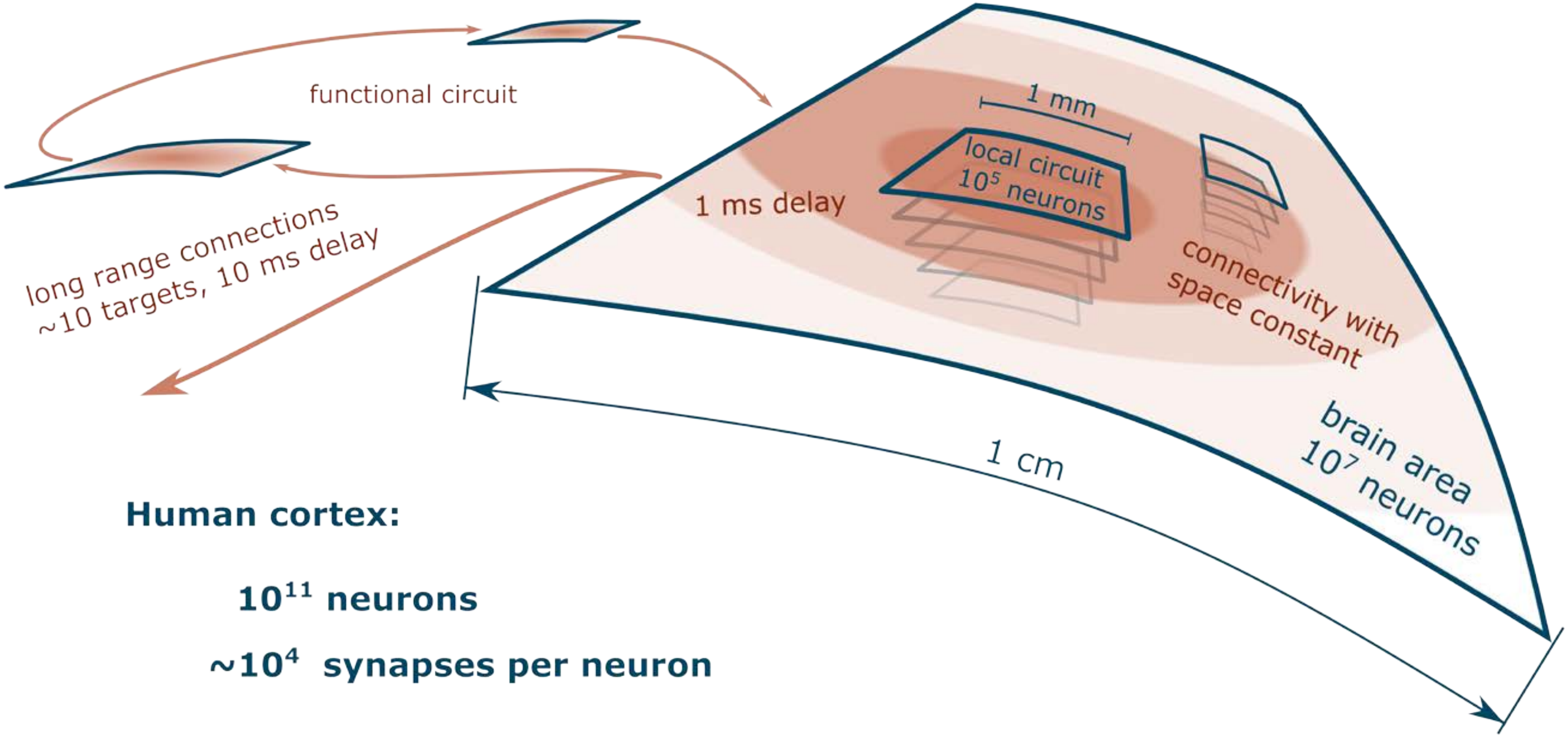
Brain-scale Connections:
- ✓ all connections
- ⇒ bottom-up and top-down approach meet on this level



Necessity of brain-scale models

- only 50% of incoming synapses of a neuron from local source
  - underconstrained model limits predictive power
  - only brain-scale model represents majority of inputs
  - self-sustained activity
- brain functions are distributed over many areas
  - brain-scale models required to close functional circuits
- requires work on software technology

Architecture of the human cortex



a network of networks with at least three levels of organization:

- connectivity of local microcircuit
- within-area connectivity with space constant
- long-range connections between areas

Meso- and macro-scale measures

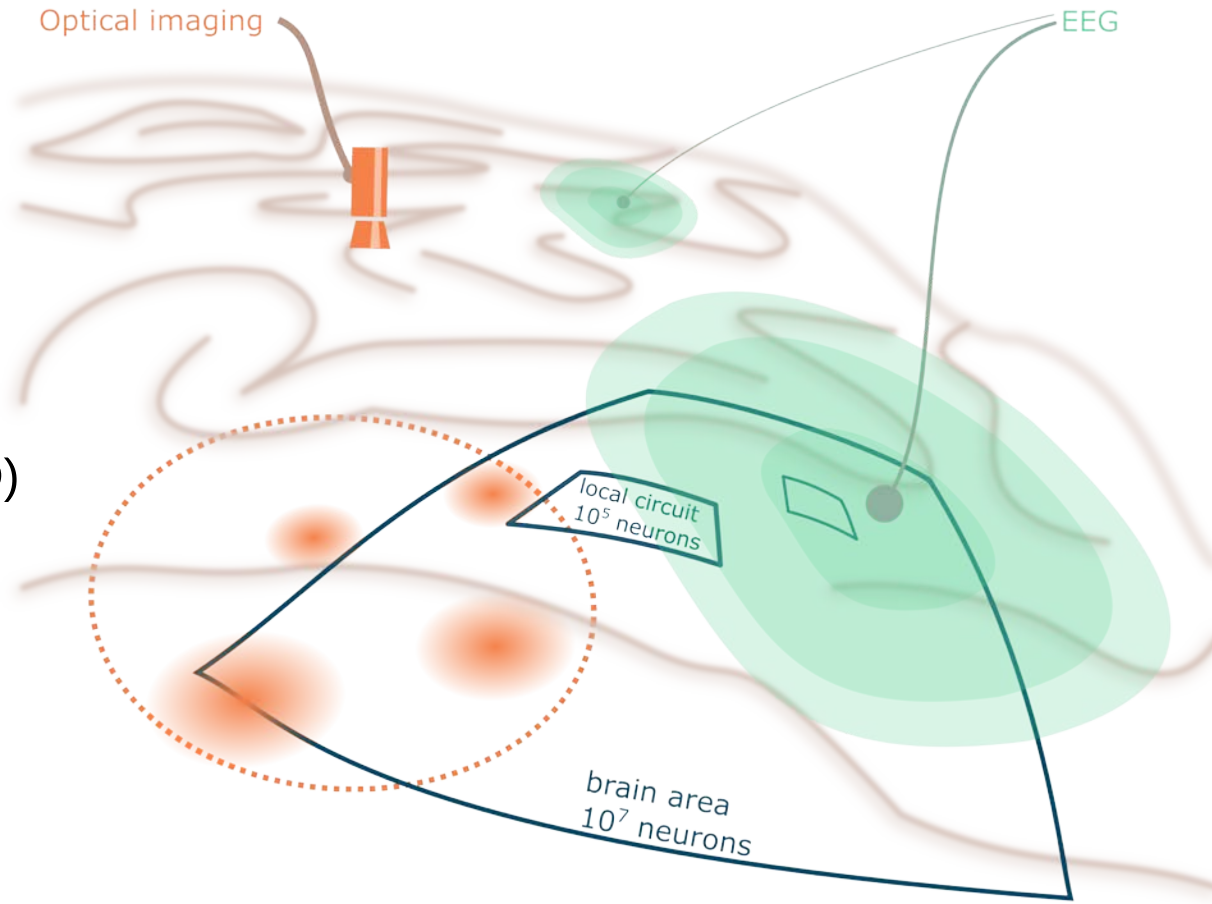
brain-scale networks
provide the substrate for

mesoscopic measures

- local field potential (LFP)
- voltage sensitive dyes (VSD)

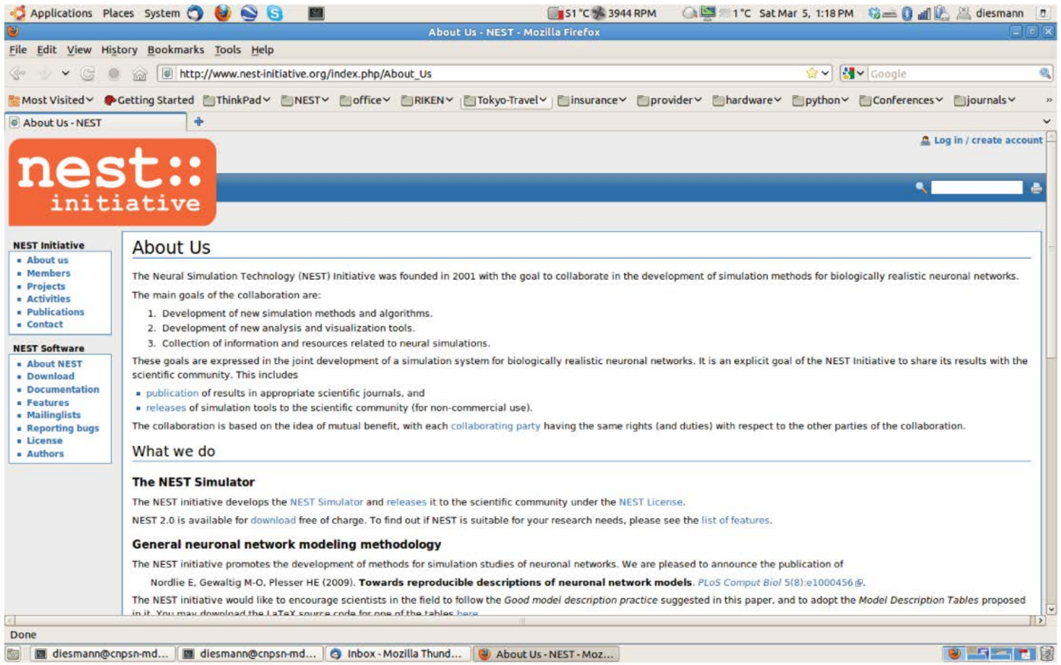
and macroscopic measures

- EEG, MEG
- fMRI resting state networks



connecting biophysical modeling to the field of neuroimaging

Collaboration



Major goals:

systematically publish new simulation technology

produce public releases under GPL

- collaboration of several labs (since 2001)
- teaching in international advanced courses
  - Okinawa Computational Neuroscience Course OCNC
  - Advanced Course in Computational Neuroscience ACCN, Europe

Scale-up to networks of 10^9 neurons

- **scale-up on K guided by 3 milestones**

1. port NEST software to K	2g*	Sep 2011	NEST 2.1
2. scale of 10^8 neurons	3g	May 2012	NEST 2.2
3. towards full brain	4g	Sep 2012	now

(*2g = 2nd generation simulation kernel)

- **scale 10^8 relevant:**
  - larger than largest area (V1)
  - enable visual cortex model respecting relative sizes
  - larger networks: long delays, sparse macroscopic connectivity

Characteristics of brain simulations

- memory overhead increases with cores
- memory not simulation time limits network size
- intention to use full memory resources: maximum-filling scaling
- analysis based on mathematical model of memory consumption:
  - Kunkel S, Potjans TC, Eppler JM, Plesser HE, Morrison A and Diesmann M (2012) Meeting the memory challenges of brain-scale network simulation. *Front Neuroinform* 5:35
  - at different scales different components of the software dominate memory consumption

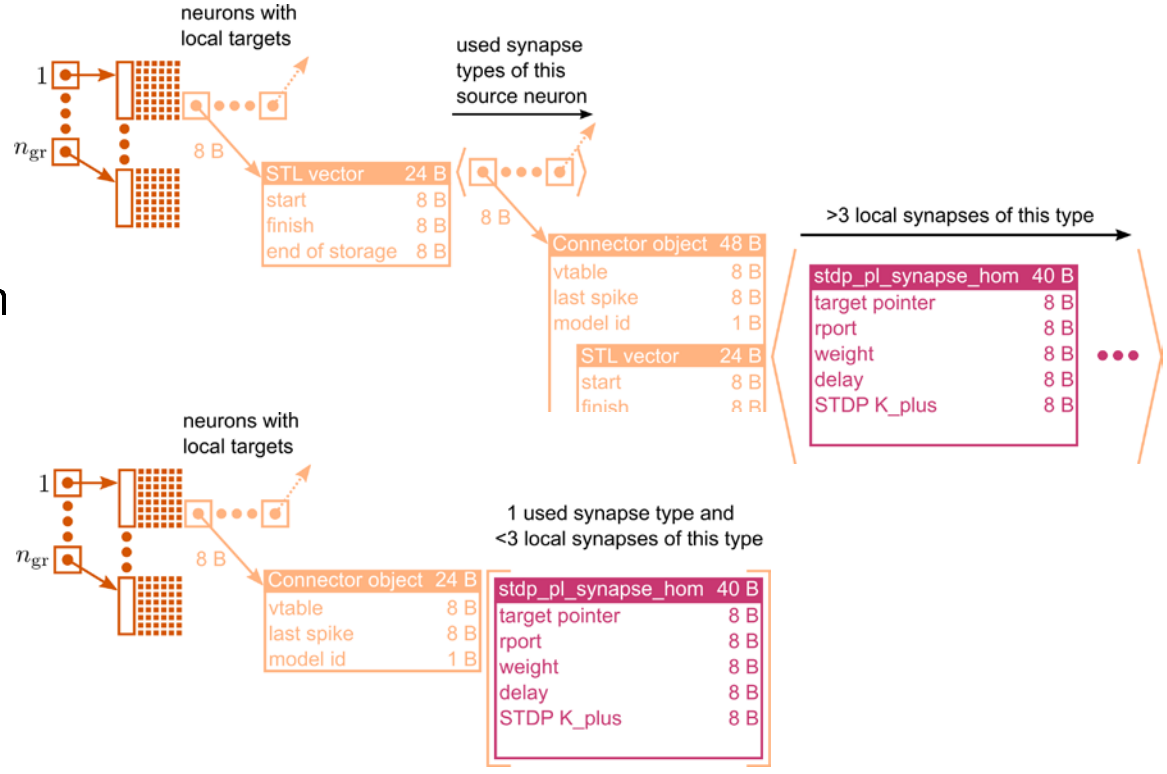
Memory layout of 3g and 4g kernel

3g memory layout

- accounts for sparseness in neuronal and connection data structures

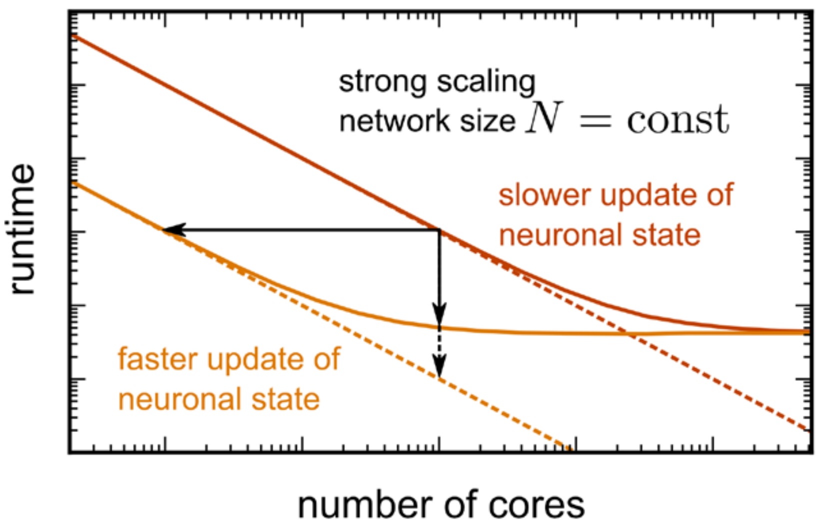
4g memory layout

- data structures account for heterogeneity of synaptic dynamics
- for > 10,000 cores, neurons with few local targets cause severe overhead
- novel adaptive data structure copes with short target lists
- not compromising on generality

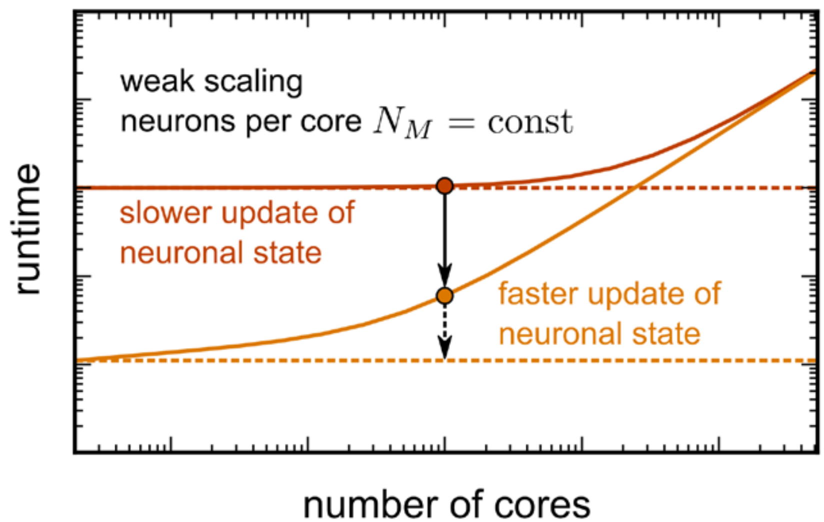


How to measure scalability

Strong scaling

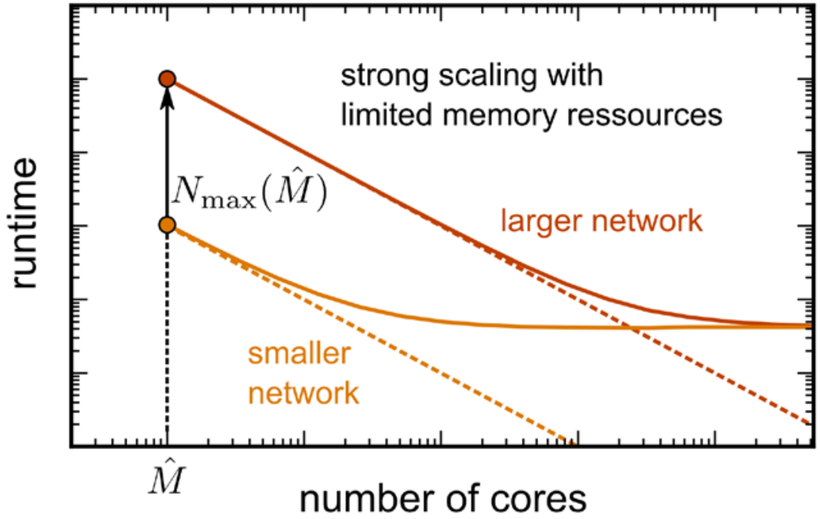


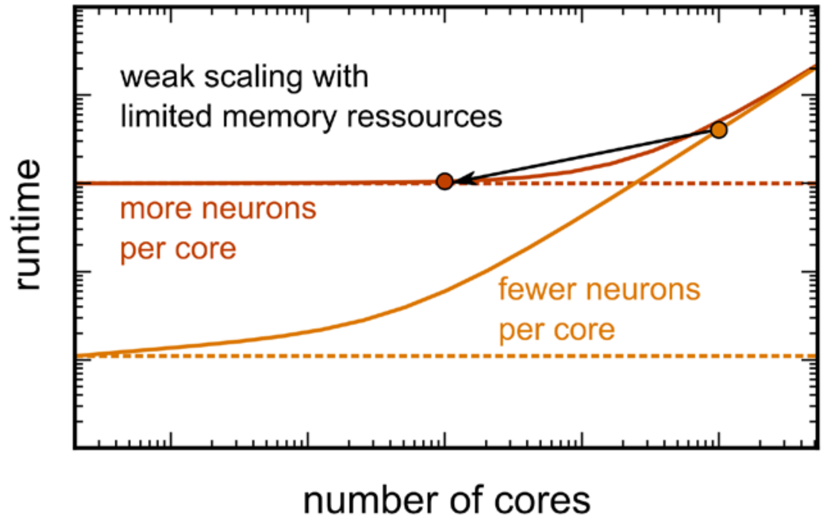
Weak scaling



- **faster element update** (decreased t_n) **leads to worse scaling** as communication dominates runtime already at fewer cores
- or the other way round: a better scaling can be achieved by using an algorithm which performs a slower element update (e.g. integration with smaller step size)

Limited memory resources





- network just fits on $\hat{M}$ cores
- improving memory consumption

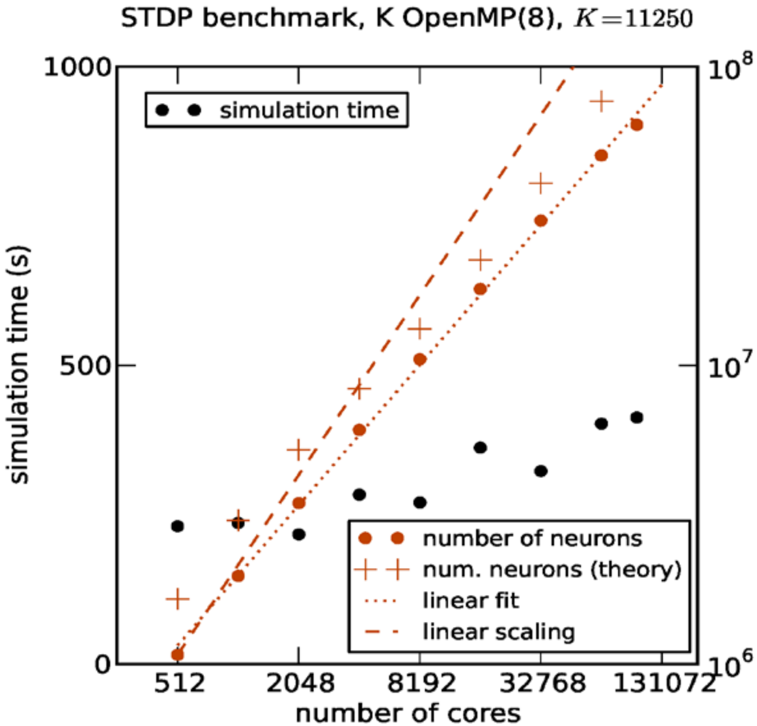
➢ larger network on $\hat{M}$ cores

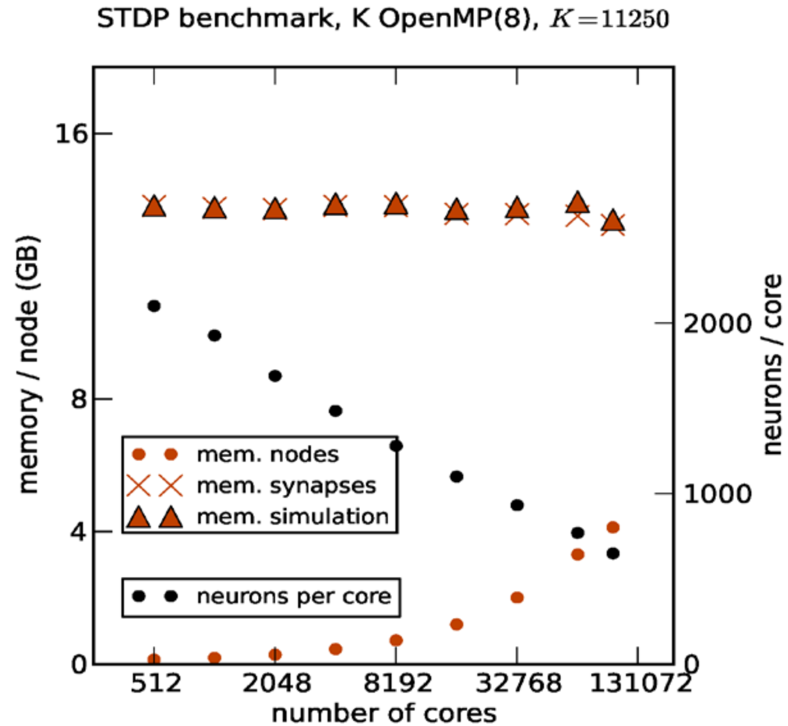
- **extreme case: same network can be simulated faster on fewer cores** (dots indicate equally sized networks)

➢ communication dominates at larger number of cores

➢ better scaling

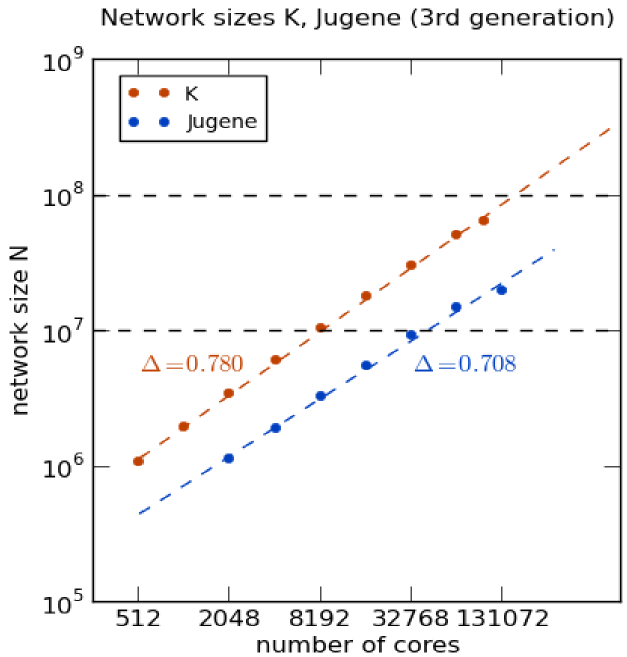
Maximum filling scaling of 3g on K

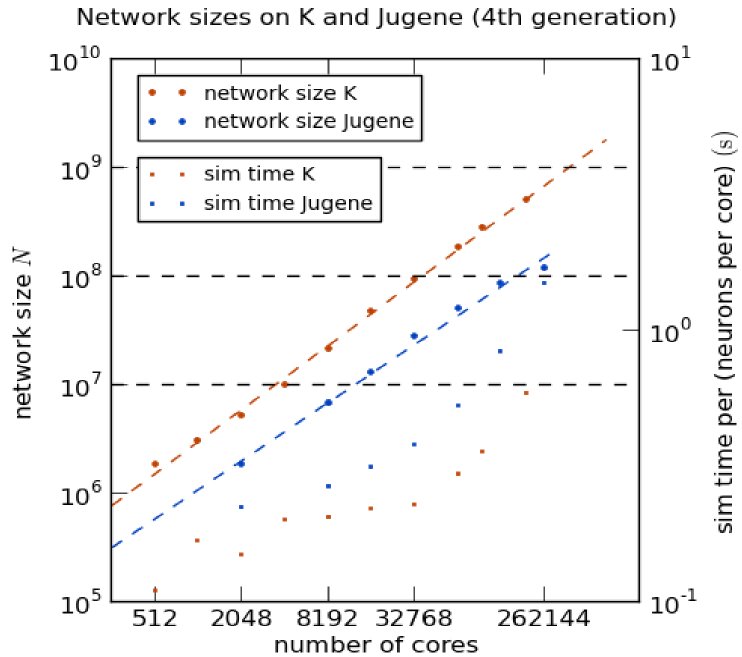




- neuroscientist interested in maximum use of resources
- hold memory close to maximum available on node (16 GB)
- neurons per core drop
- simulation time increases due to increased communication

Performance of 3g and 4g on Jugene and K





- **end of dashed lines indicate machine size, the potential of K**
- JUGENE largest number of MPI jobs tested 65,536 = 262,144 cores
- 74% of MPI jobs required for full K tested
- K largest number of MPI jobs tested 24,576 = 196,608 cores
- wall clock time 1 biological second of the largest 4g simulation on K: 1520s = 25min

Publication on technology of 3g kernel

frontiers Journals Events Jobs People Videos News Blogs

About Submit Register Login Search

frontiers IN NEUROINFORMATICS

Info
- Home
- About
- Editorial Board
- Archive
- Research Topics
- View Some Authors
- Review Guidelines
- Subscribe to Alerts

Search

Article Type: All

Publication Date

ORIGINAL RESEARCH ARTICLE

Share 0 Like 0 Comment 0 29 Share Altmetric 0 285 Views

Front. Neuroinform., 02 November 2012 | doi: 10.3389/fninf.2012.00026

Supercomputers ready for use as discovery machines for neuroscience

Moritz Helias[1,2*], Susanne Kunkel[1,3,4], Gen Masumoto[5], Jun Igarashi[6], Jochen Martin Eppler[1], Shin Ishii[7], Tomoki Fukai[6], Abigail Morrison[1,3,4,8] and Markus Diesmann[1,2,4,9]

1 Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Jülich Research Centre, Jülich, Germany
2 RIKEN Brain Science Institute, Wako, Japan
3 Simulation Laboratory Neuroscience - Bernstein Facility for Simulation and Database Technology, Institute for Advanced Simulation, Jülich Aachen Research Alliance, Jülich Research Centre, Jülich, Germany
4 Bernstein Center Freiburg, Albert-Ludwig University of Freiburg, Freiburg, Germany
5 High-Performance Computing Team, RIKEN Computational Science Research Program, Kobe,

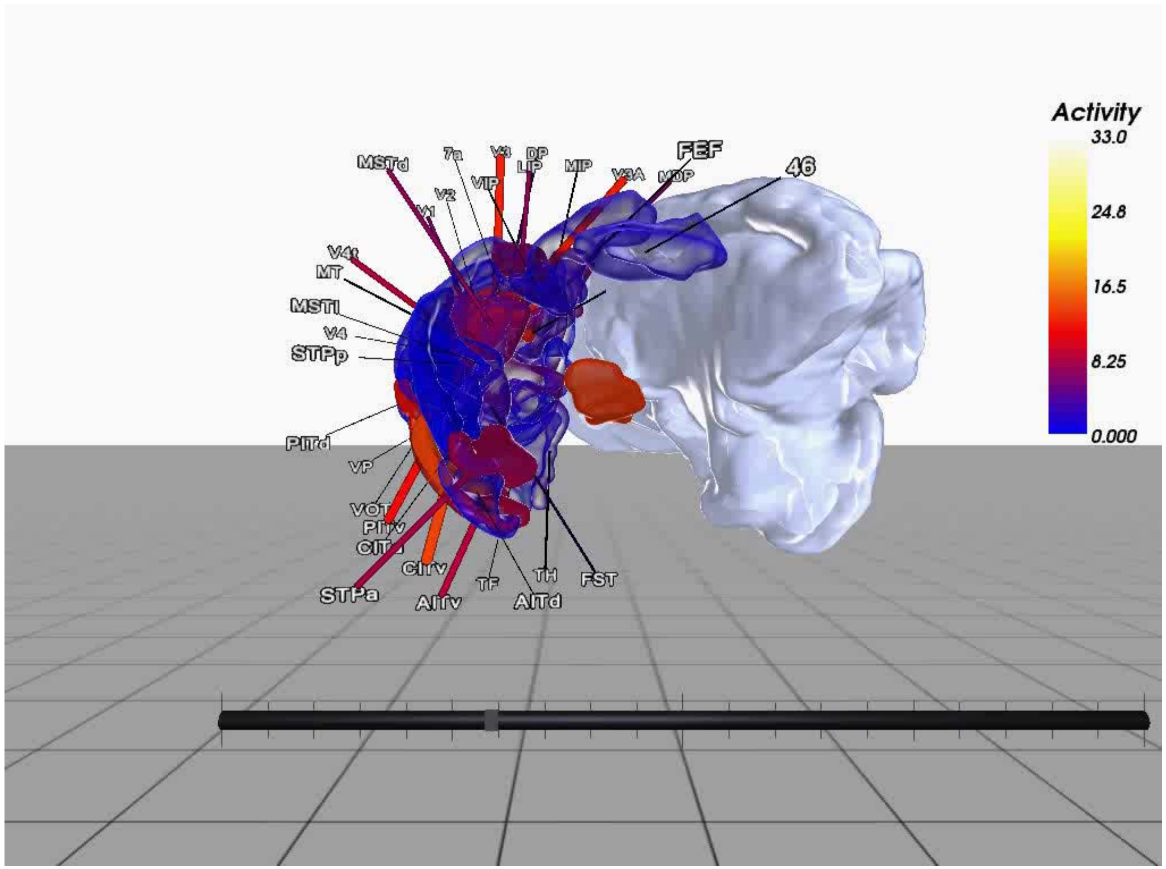
Article Info
- Abstract
- Full Text
- PDF
- Export Citation
- XML
- View Article Impact

The Authors in
- Frontiers
- Google
- Google Scholar
- PubMed

Related Article
- in Frontiers

Mitglied in der Helmholtz-Gemeinschaft

Visualization



- Virtual Reality Group, Torsten Kuhlen, RWTH Aachen University

I/O characteristics of NEST

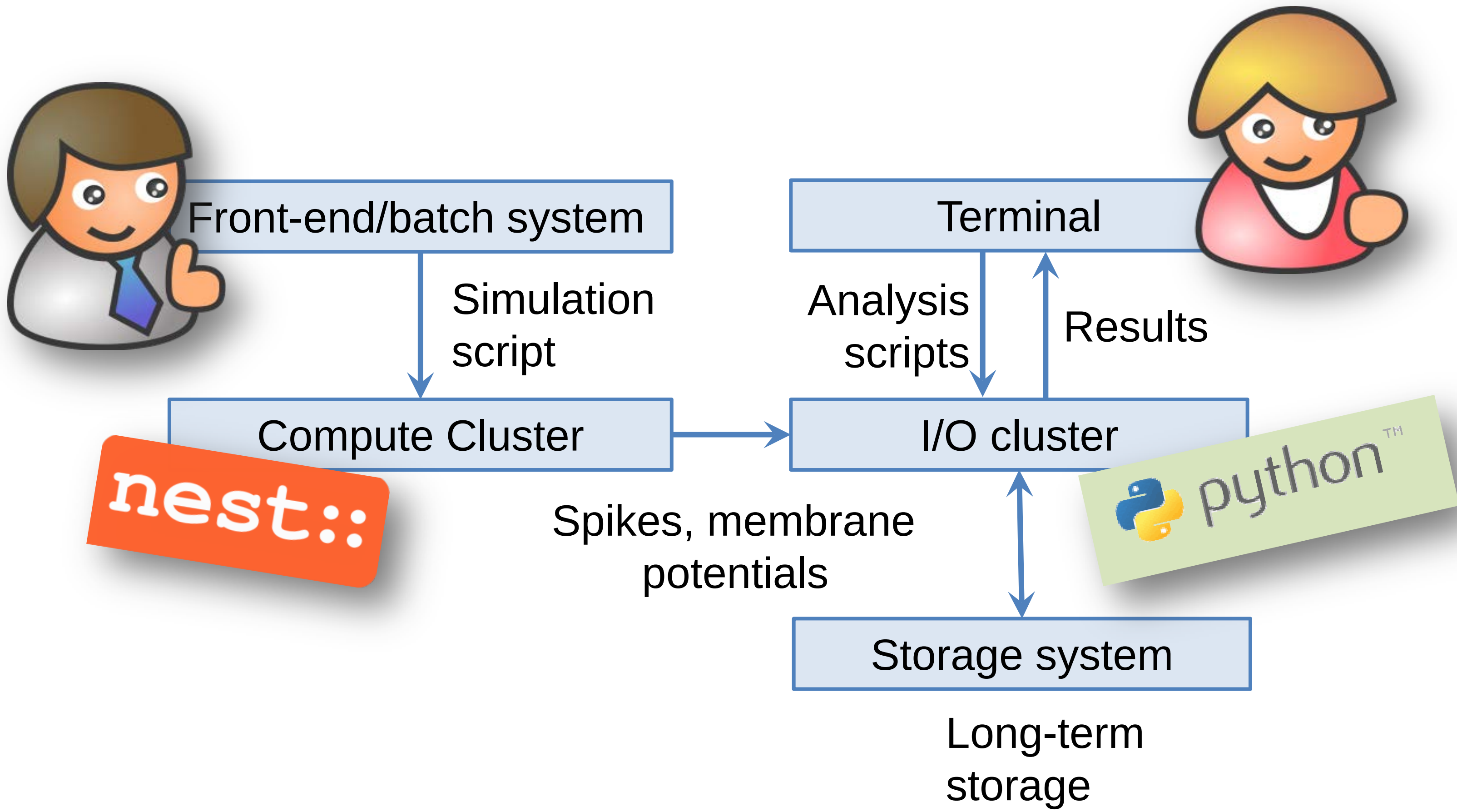
- currently no state-of-the-art simulator can write out data!

- we need to write small chunks of data continuously
- all processes write the data generated locally

- 10^8 neurons (each spiking at ~1/s) means 1.49 GB per simulated s
- recording membrane potentials once every 10ms means 149 GB
- recording synaptic weights multiplies by ~10^4

- in practice we only need data by a subset of neurons/synapses!

Our planned use of BGAS



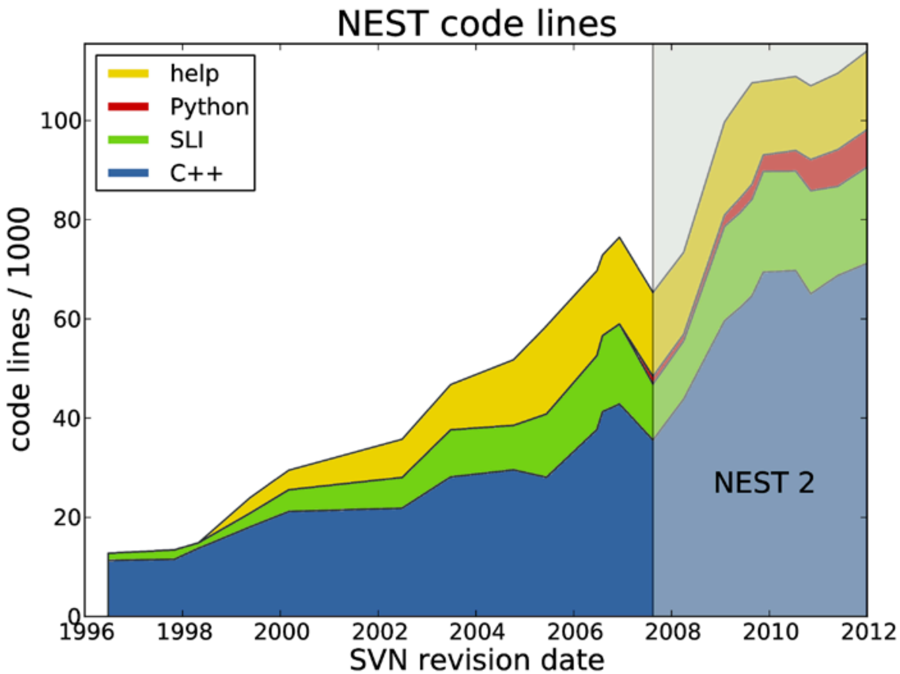
Software challenges to use BGAS

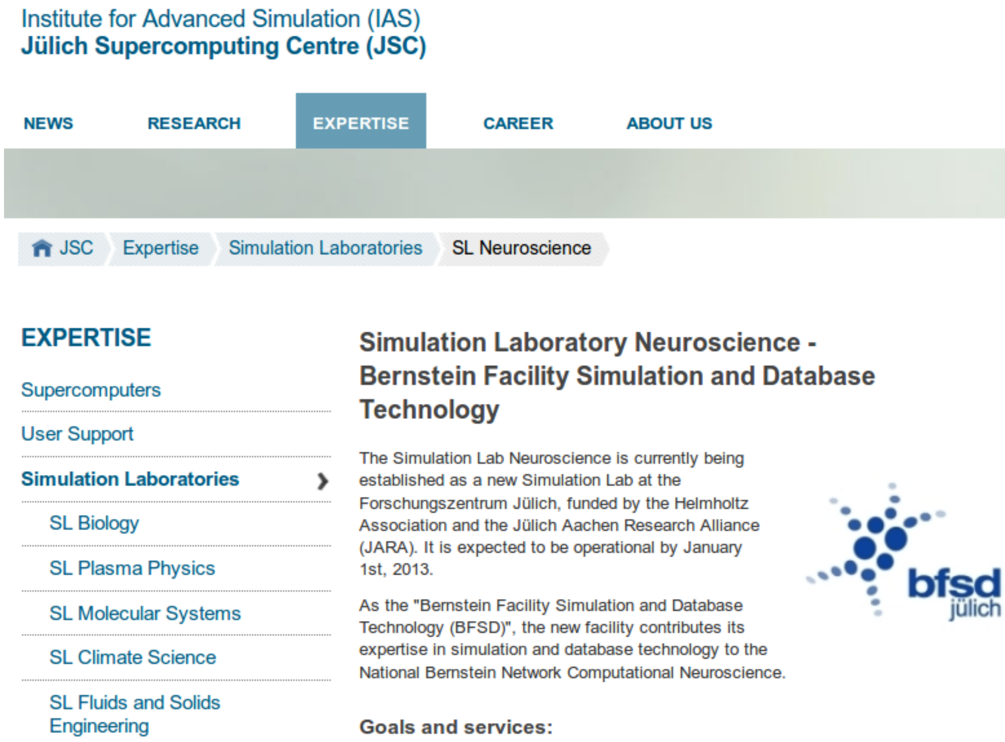
- Currently, data analysis is carried out after the simulation is over

- To use BGAS, we need to intertwine simulation and analysis
- Data might need to be collected before writing to minimize the amount of files created on i/o nodes

- We have a prototype implementation for parallel i/o based on SIONlib (with group of Dirk Pleiter)
- we require a high degree of portability; success of NEST is based on the fact that it runs on laptops and supercomputers

Complexity of software development





- complexity requires continuous adaptation of new technologies

Summary

- model of local cortical network explains basic dynamical properties
- remains underconstrained because only 50% of connections are local
- functional loops only closed at the brain-scale
- 10^8 milestone achievable with 3g kernel (milestone 2)
- 10^9 barrier in sight with 4g kernel (milestone 3)
- no compromise on generality
- short run times enable use of K as discovery machine for neuroscience

- next steps:
    - deliver 4g kernel to our partners
    - concepts for exa-scale machines

Acknowledgements

Moritz Helias

Susanne Kunkel

Jochen Martin Eppler

Hans Ekkehard Plesser

Marc-Oliver Gewaltig

Abigail Morrison

Shin Ishii

Tomoki Fukai

Jun Igarashi

Makoto Taiji

Gen Masumoto

Yousuke Ohno

K, JUGENE, JUQUEEN

	K	JUGENE	JUQUEEN
CPU clock speed	2.0 GHz	850 MHz	1.6 GHz
Cores / compute node	8	4	16
Hardware threads / core	1	1	4
Memory / core	2 GB	500 MB	1 GB
Total nodes	88,128 (864 racks)	73,728 (72 racks)	24,576 (24 racks)
Total cores	705,024	294,912	393,216
Total hardware threads	705,024	294,912	1,572,864
Total memory	1,377 TB	144 TB	384 TB

How to measure scalability

Simple model of runtime

$$T = \underbrace{N_M t_\mathrm{n}}_{\text{update of local neurons}} + \underbrace{\log_2(M) t_s + (M-1) t_\mathrm{w} m}_{\text{MPI Allgather}}$$

$N_M = \frac{N}{M}$	number of neurons per MPI process
N	total number of neurons in the network
M	number of MPI processes
t_s	start-up time
t_w	time to transfer 1 Byte
m	message size in Bytes
t_n	time needed to update one neuron