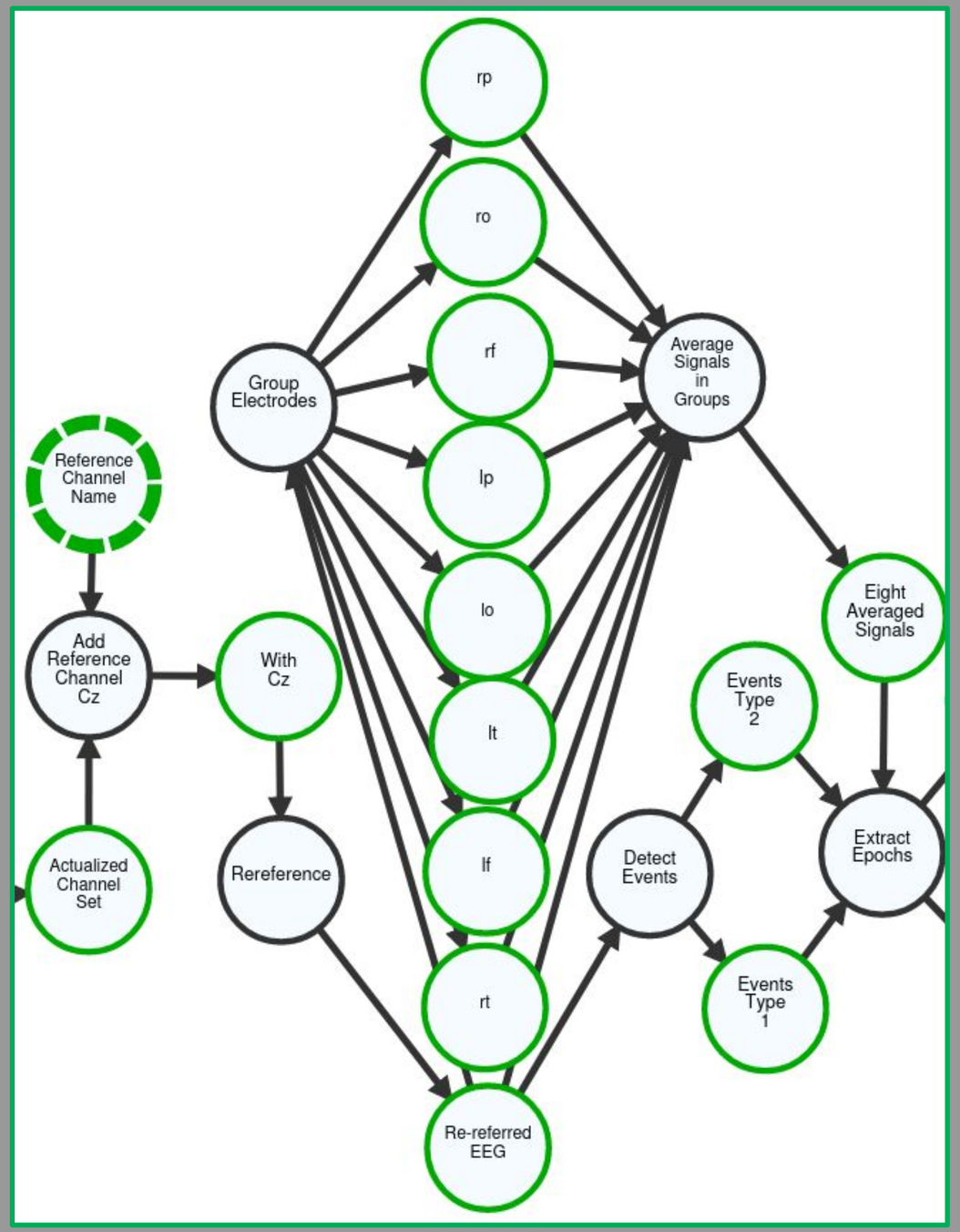


EDM

Automation of EEG Data Processing with HPC Community Cloud

Maxim A. Gorodnichev[αβγδ]

Elizaveta D. Nalepova[β]
Ekaterina A. Merkulova[αβ]
Pavel D. Rudych[β]
Alexandr N. Savostyanov[αβ]

[α] Scientific Research Institute of Neurosciences and Medicine
[β] Novosibirsk State University
[γ] Institute of Computational Mathematics and Mathematical Geophysics SB RAS
[δ] Novosibirsk State Technical University

Outline

1. The Problem
2. The Objective
3. Related Works
4. Computational Models
5. EEG Processing: Computational Model Example
6. Concept: Interactive Supercomputer Application
7. Concept: Mini-Applications
8. Implementation Framework
9. Summary

The problem: target audience

1 Target Audience

- neuroscientists
- generally: those who are not satisfied with existing data processing instruments and combine data processing / simulation tools into own pipelines

The problem: routine activities

1 Target Audience

- neuroscientists
- generally: those who are not satisfied with existing data processing instruments and combine data processing / simulation tools into own pipelines

2 Researchers are involved in the following activities

- data gathering, data source management
- data processing pipeline composition
- data movement, running pipelines on high performance computing systems
- data presentation/visualization
- providing reproducibility, publications
- building applications for end users

The problem: what is missing?

1 Target Audience

- neuroscientists
- generally: those who are not satisfied with existing data processing instruments and combine data processing / simulation tools into own pipelines

2 Researchers are involved in the following activities

- data gathering, data source management
- data processing pipeline composition
- data movement, running pipelines on high performance computing systems
- data presentation/visualization
- providing reproducibility, publications
- building applications for end users

3 The Problem

- No high level tools are available to support these activities in a systematic way

The problem: an approach

1 Target Audience

- neuroscientists
- generally: those who are not satisfied with existing data processing instruments and combine data processing / simulation tools into own pipelines

2 Researchers are involved in the following activities

- data gathering, data source management
- data processing pipeline composition
- data movement, running pipelines on high performance computing systems
- data presentation/visualization
- providing reproducibility, publications
- building applications for end users

3 The Problem

- No high level tools are available to support these activities in a systematic way

4 Required:

- a general methodical approach
- a software platform/framework to implement the approach

The objective

- Develop a method and a software framework for formal specification and implementation of data processing pipelines on supercomputers in neurophysiology.

The objective

- Develop a method and a software framework for formal specification and implementation of data processing pipelines on supercomputers in neurophysiology.

The particular scientific interest and the novelty of this project are in the study of **computational models** as a formal tool for representation and management of knowledge about computations in an application domain.

Related Works

- “Tool Boxes”:
  - EEGLAB
  - Python-MNE
- Workflow management software
  - Apache Airflow
  - Apache Beam
  - Apache Hadoop + Hue
  - ActiveEon
  - Prefect
- Research Data Management
  - DataLad
  - iRODS
- Domain-Specific Language Development
  - Delite
  - AnyDSL

Related Works

- “Tool Boxes”:
  - EEGLAB
  - Python-MNE
- Workflow management software
  - Apache Airflow
  - Apache Beam
  - Apache Hadoop + Hue
  - ActiveEon
  - Prefect
- Research Data Management
  - DataLad
  - iRODS
- Domain-Specific Language Development
  - Delite
  - AnyDSL

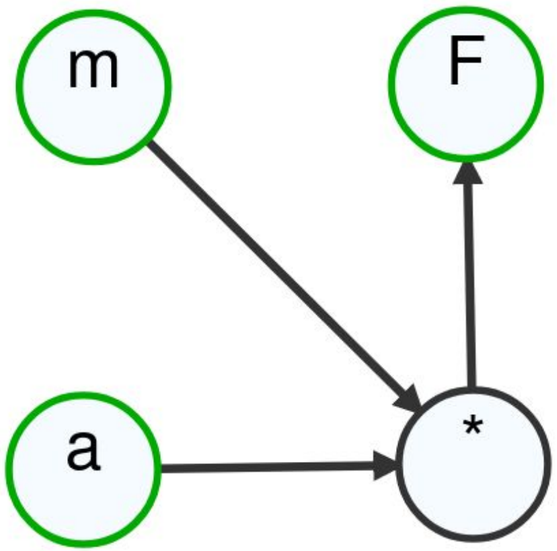
What is missing? :

A systematic approach to accumulation and automatic reuse of domain-specific knowledge about computations

Computational Model of a Problem Domain (PD)

– a formal specification of knowledge about computations in PD:

- **variables** : represent objects/entities/quantities in a PD
- **operations** : connect variables



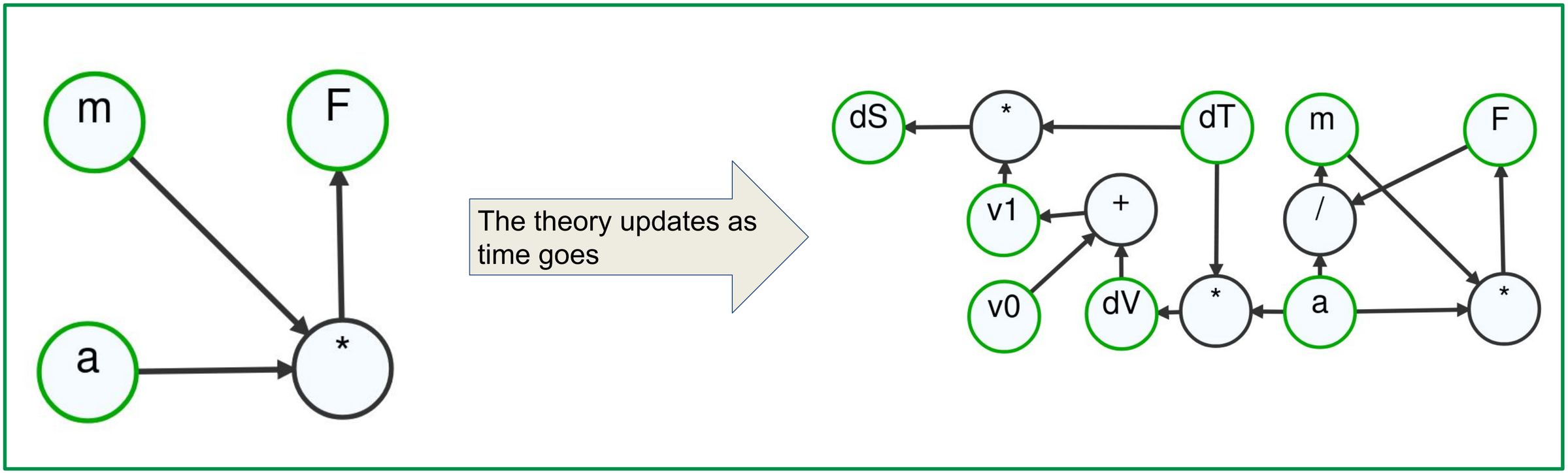
Example: CM in Mechanics

Computational Model of a Problem Domain (PD)

– a formal specification of knowledge about computations in PD:

variables : represent objects/entities/quantities in a PD

operations : connect variables

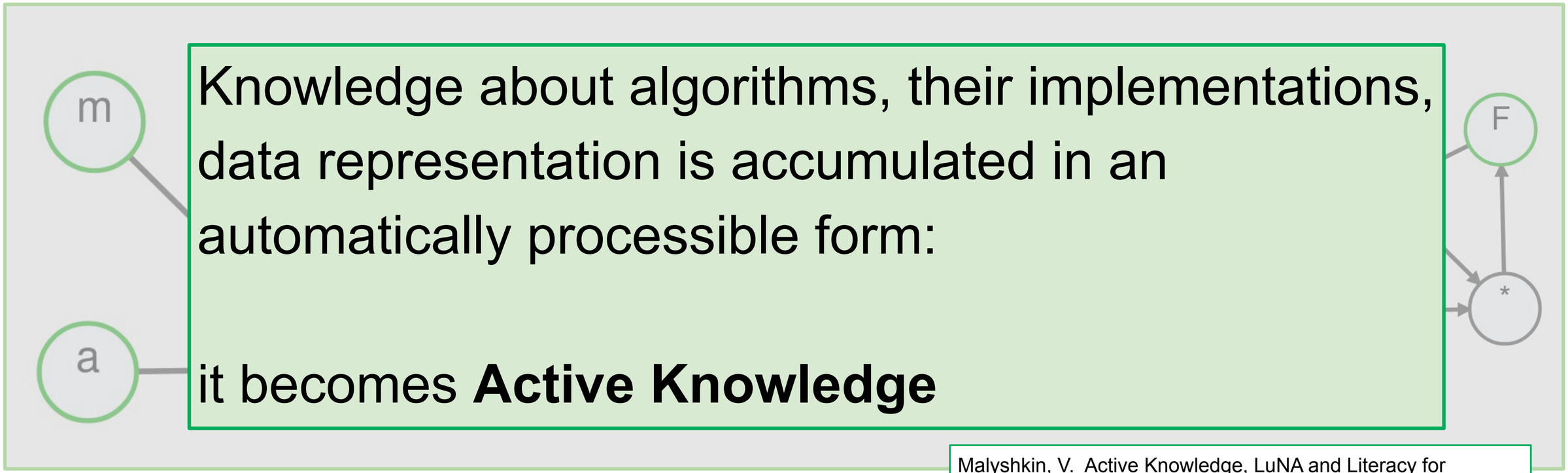


Example: CM in Mechanics

Computational Model of a Problem Domain (PD)

– a formal specification of knowledge about computations in PD:

- **variables** : represent objects/entities/quantities in a PD
- **operations** : connect variables



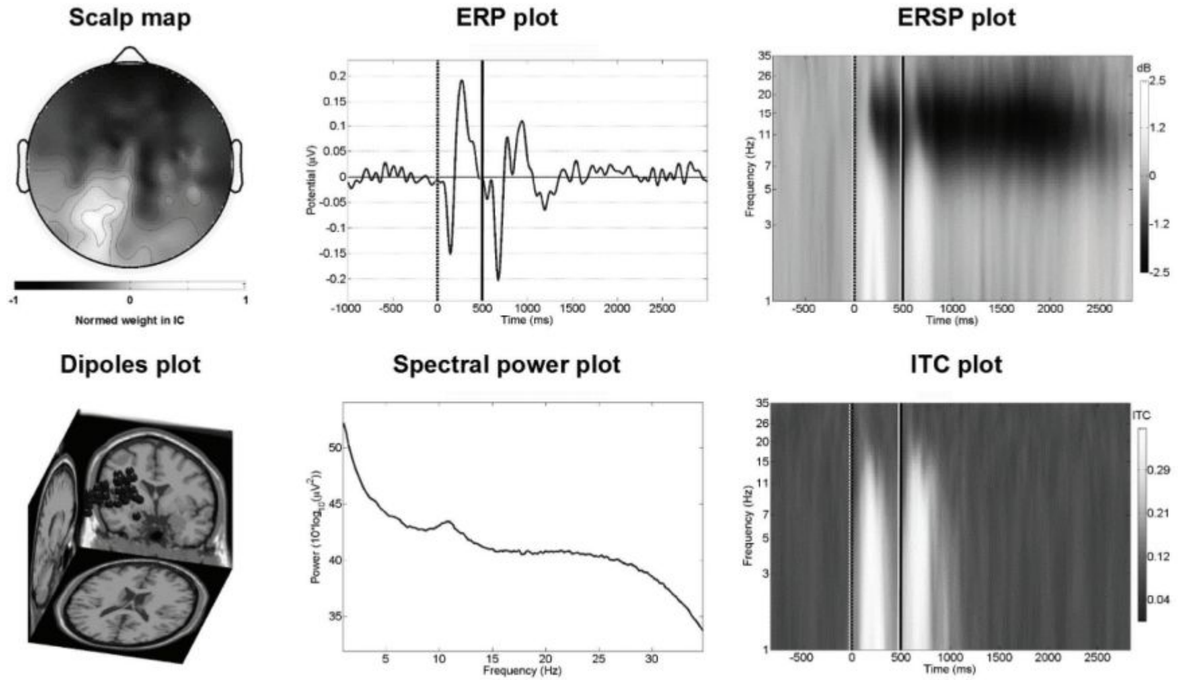
Example: CM in Mechanics

Malyshkin, V. Active Knowledge, LuNA and Literacy for Oncoming Centuries, doi: 10.1007/978-3-319-25527-9_19

Problem Domain: Encephalography Data Processing

- EEG is a method of noninvasive exploration of functional brain activity, its applications:
  - scientific studies of neurocognitive processes,
  - diagnosis,
  - treatment and prognosis of some diseases

- It records electric potentials generated by neurons by reading the signal from electrodes placed over a head.



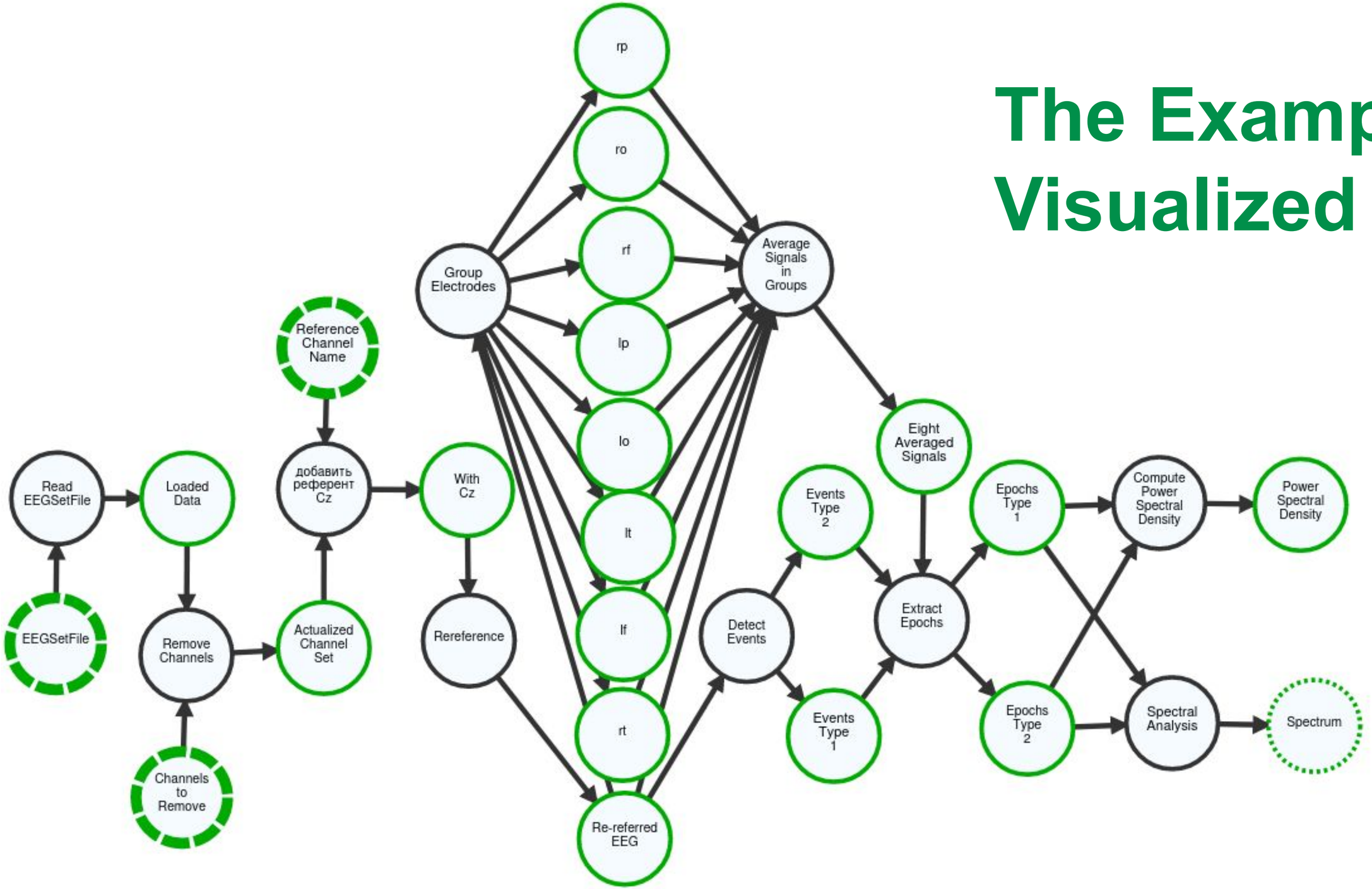
Analysis of EEG records

doi:10.4018/ijehmc.2014040103

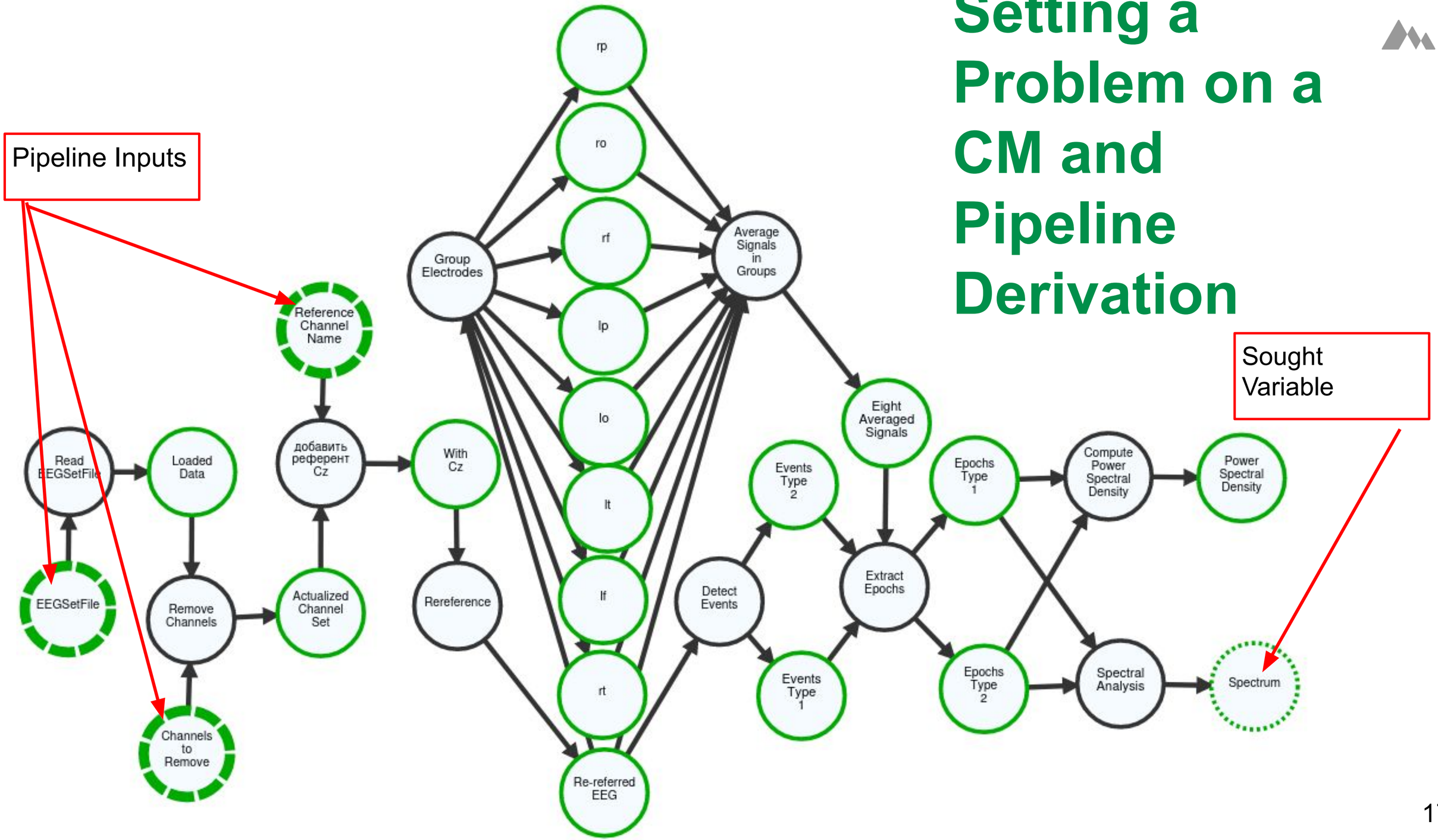
Example of an EEG Processing Computational Model

Operation	Inputs	Outputs
Read an EEG file	• EEG file	• Loaded EEG
Remove Channels	• Loaded EEG, • List of Channels to Remove	• Actual EEG Channels
Add a Reference Channel	• Actual EEG Channels, • Reference Channel Name	• Actual EEG Channels with a Reference
Re-refer (re-refers EEG to an average)	• Actual EEG Channels with a Reference	• Re-referred EEG
Divide Channels into 8 Lobe Groups	no	8 groups of channels (each group contains channel IDs), left/right temporal, frontal, occipital, parietal: • lt, rt, lf, rf, lo, ro, lp, rp
Average Signals in Groups	• Re-referred EEG, • lt, rt, lf, rf, lo, ro, lp, rp	• Group of Averaged Signals
Detect Events	• Re-referred EEG	arrays that specify events in the Re-referred EEG: • Events Type 1 (eyes closed), • Events Type 2 (eyes opened)
Extract Epochs	• Events Type 1, • Events Type 2, • Group of Averaged Signals	• Epochs Type 1 • Epochs Type 2
Compute Power Spectral Density	• Epochs Type 1 • Epochs Type 2	Power Spectral Density (per epoch, per epoch type)

The Example Visualized



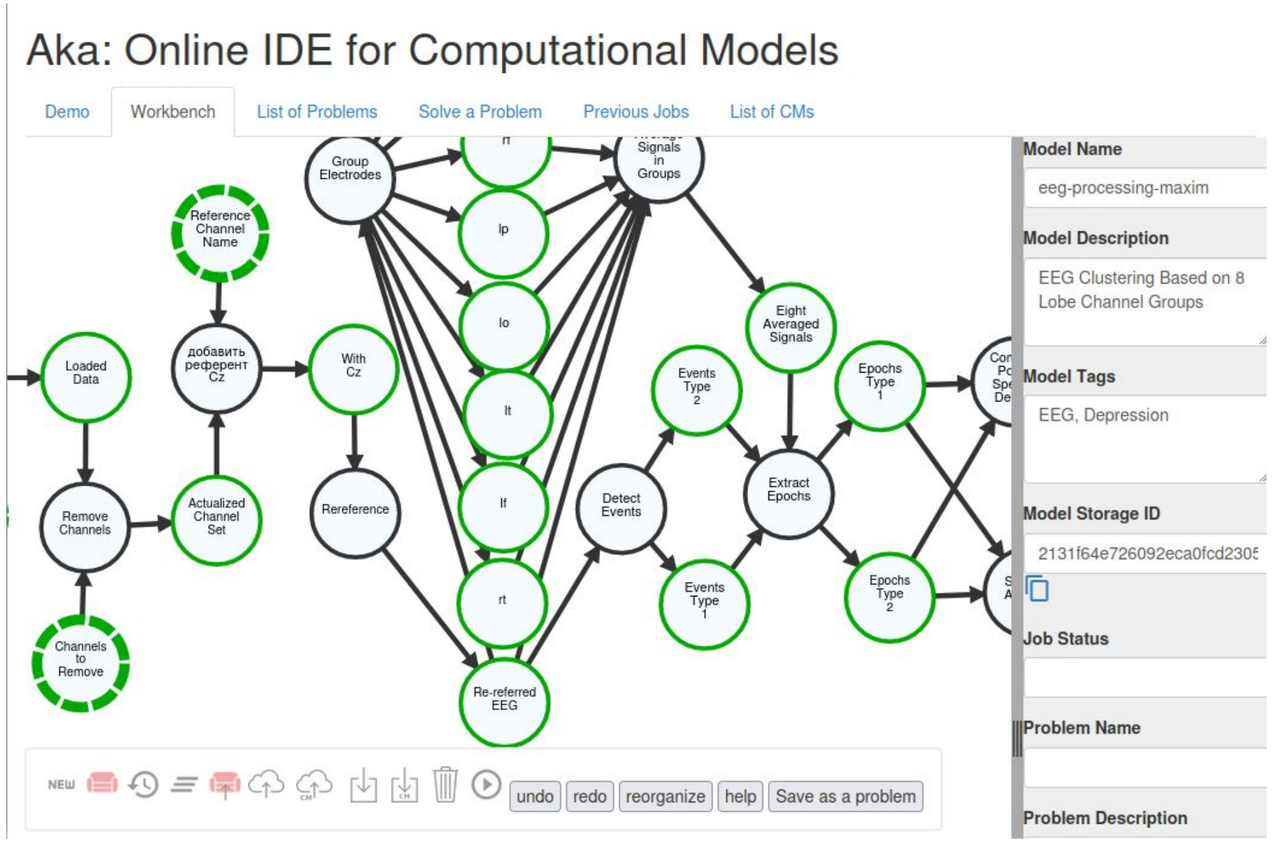
Setting a Problem on a CM and Pipeline Derivation



Concept: Interactive Supercomputer Application

ISA consists of:

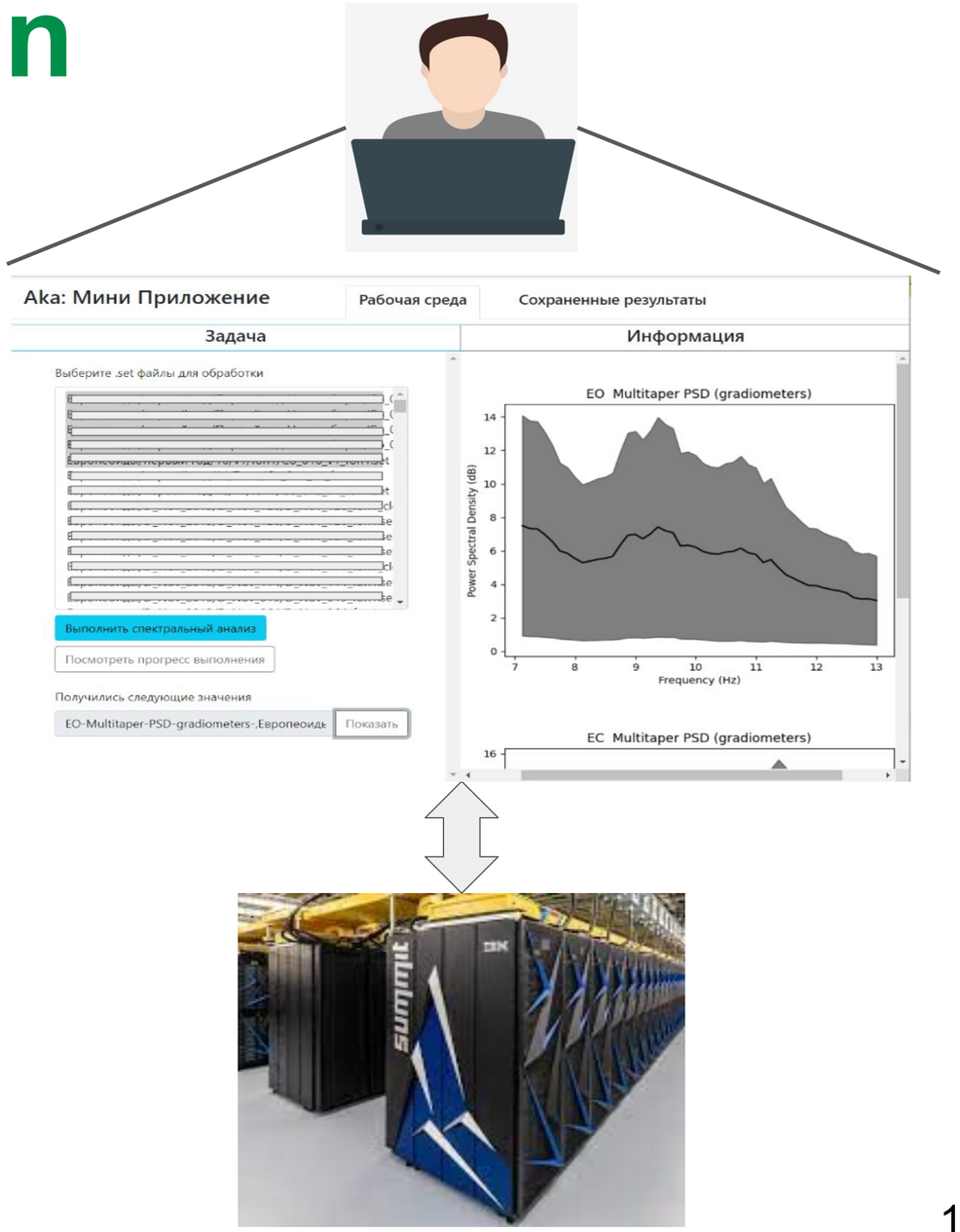
- a **workbench** to develop CMs and work with them in a general way
- extendable collection of generated GUI-enabled **mini-applications** that implement specific (derived) pipelines



CM Workbench

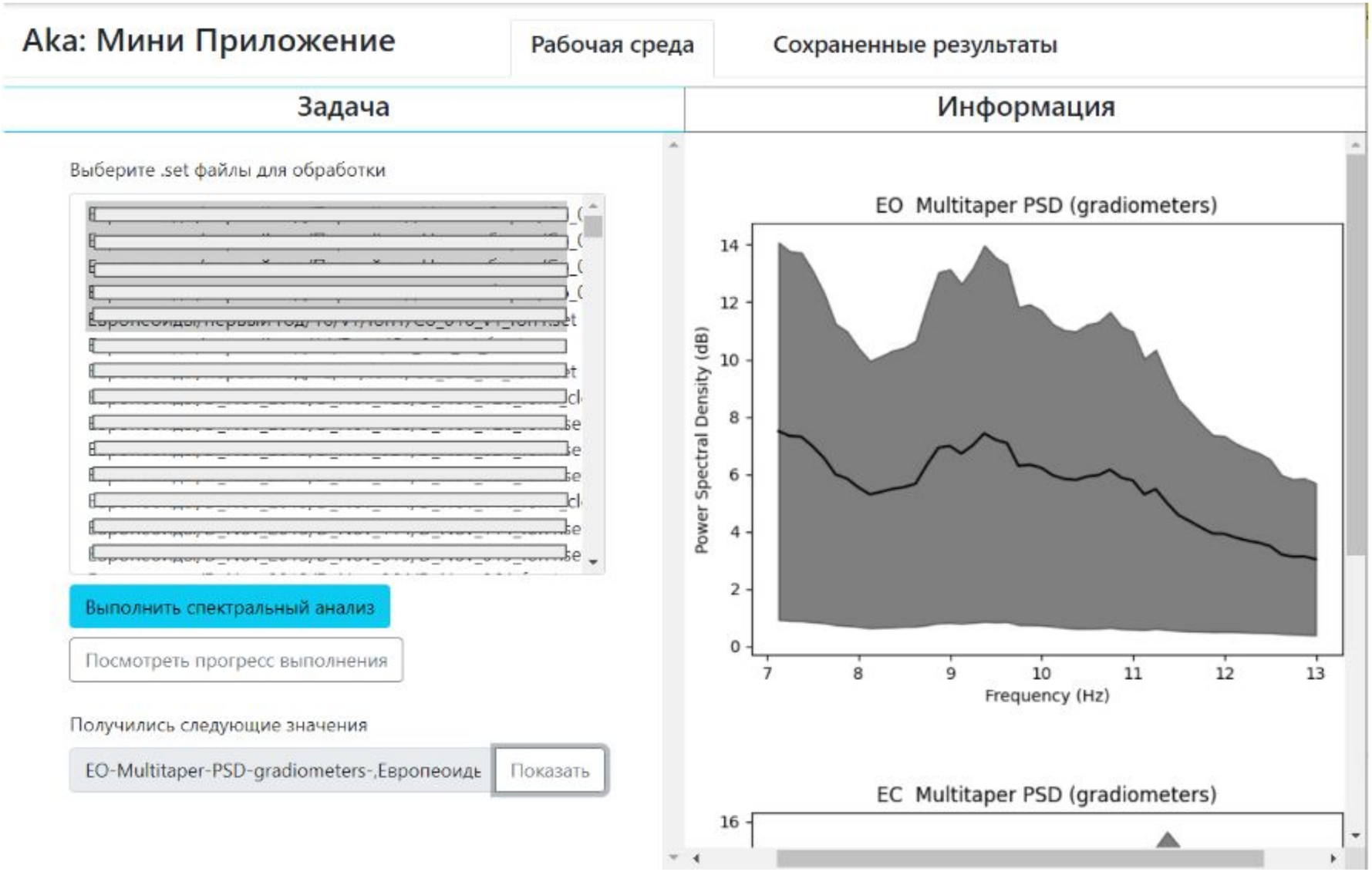
Concept: Mini-Application

- is intended for end-users
- provides GUI for a specific pipeline allowing a user to:
  - set inputs,
  - start processing,
  - monitor processing,
  - see the results.
- can be automatically generated based on
  - a computational model,
  - a library of widgets to represent variables,
  - a library of operations implementations

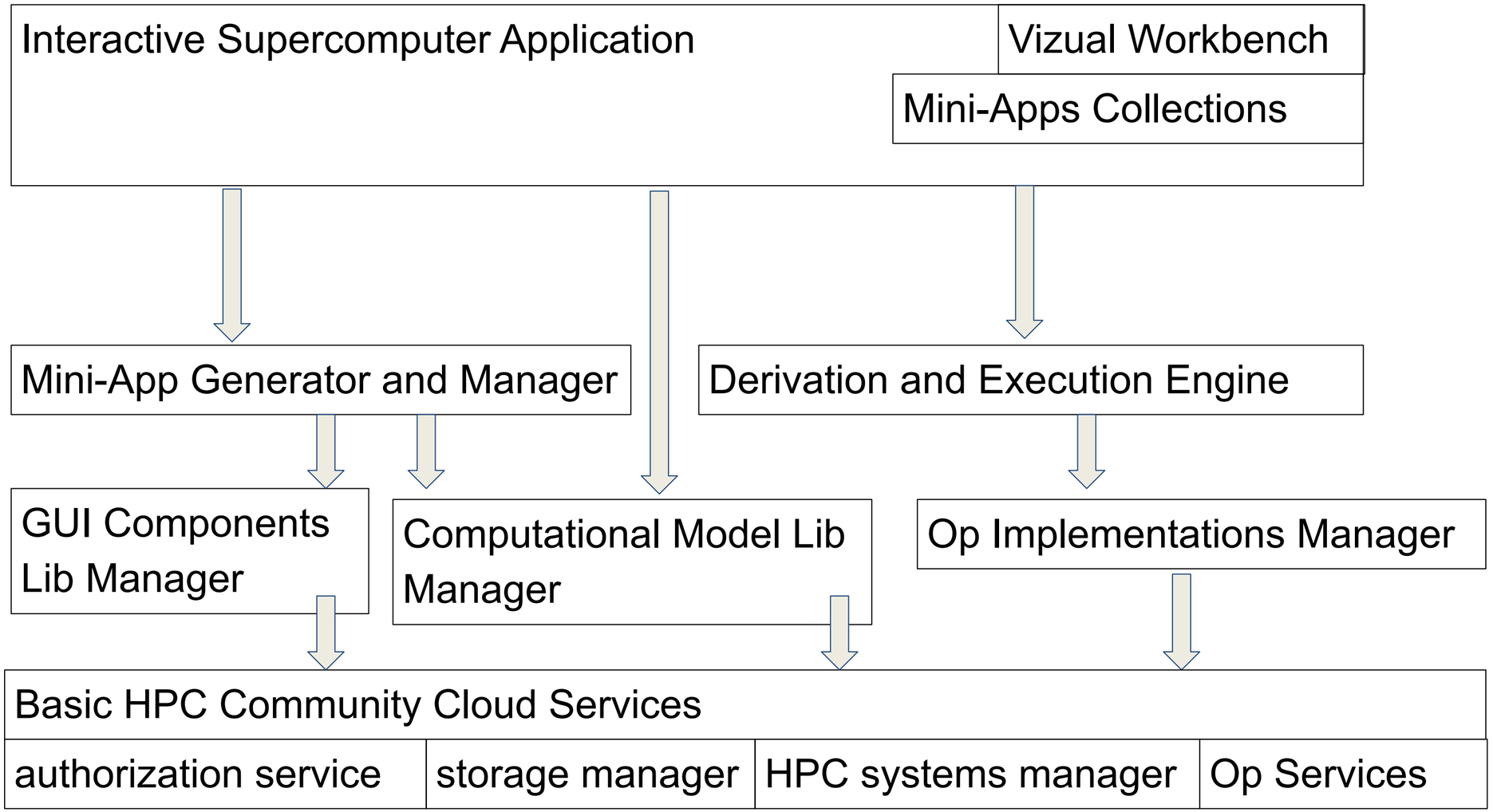


Mini-apps can be organized into specific toolboxes (application suits) for particular user groups

Mini-Application Example



HPC Community Cloud: General Framework Architecture



Summary

- An approach based on computational models formalism is proposed to systematically organize activities associated with neurophysiology data processing
- An example computational model is proposed to study implementation problems of the approach
- A prototype implementation of a framework is done as a part of the HPC Community Cloud project

Future work:

- development of computational models and applications to better understand requirements
- addressing a wide spectrum of problems in efficient implementation of all framework components

Acknowledgements

- The study was supported by the Russian Science Foundation (RSF)

  No 22-25-00735

- The Siberian Branch of the Russian Academy of Sciences (SB RAS) Siberian Supercomputer Center provided supercomputer facilities.