*10th International Conference*
*"Distributed Computing and GRID Technologies in Science and Education"*

# *Supercomputing Co-Design: No Other Options*


*Prof. Vladimir Voevodin*
*Research Computing Center, MSU, Director*
*MSU Branch in Sarov, Director*
*CMC Faculty, MSU*

*voevodin@parallel.ru*


Joint Institute for Nuclear Research
Meshcheryakov Laboratory of Information Technologies
10 GRID2023
3-7 July 2023
10th International Conference
"Distributed Computing and Grid Technologies in Science and Education"

*July, 3rd, 2023, JINR, Dubna*

# *Supercomputing technologies are everywhere today*


Supercomputer technologies are an efficient way to ensure competitiveness of science, companies, industry, economy of countries.
Produce faster, cheaper, better, bring to market faster…

# *Moscow University Supercomputing Facilities*


Lomonosov-2 supercomputer of Moscow University is a key element of the national supercomputing infrastructure for science and education in Russia.

# *Computational fluid dynamics of the smallest insects*

*Polilov A.A., Farisenkov S.E., Lapina N.A. Lomonosov Moscow State University,*
*Kolomensky D.S. Skolkovo Institute of Science and Technology*

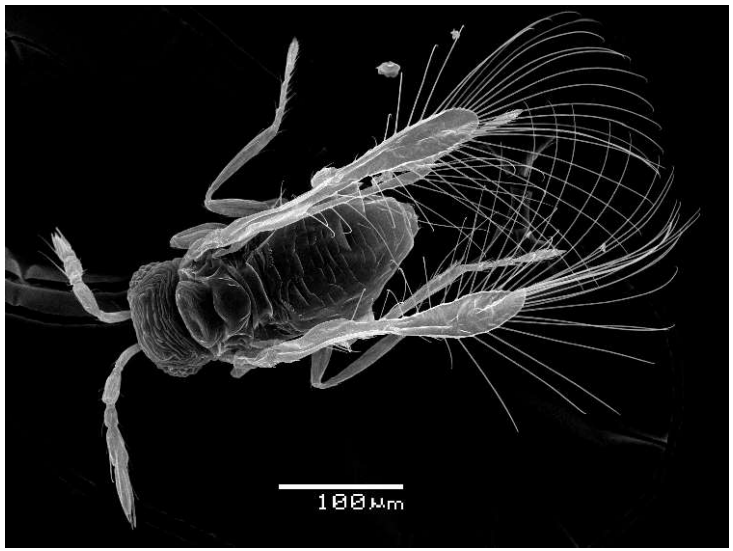
DRIVER. *Calculate the aerodynamic forces exerted on the smallest flying insects with body length less than 0.5 mm.*

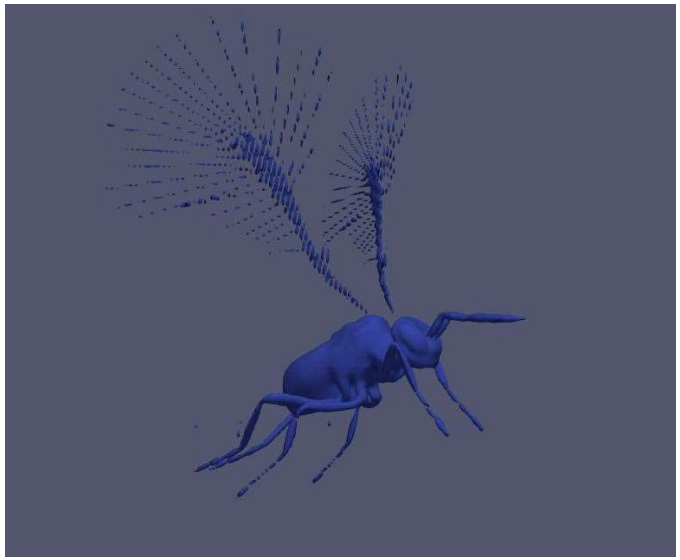
STRATEGY. *Numerical solutions of the Navier-Stokes equation using a finite-difference method with dynamic grid adaptation and 3D geometrical models obtained from electron scanning microscope data.*

OBJECTIVE. *Analyze the aerodynamic properties of bristled wings of the smallest insects.*

IMPACT. *Understanding of the mechanical function of flight of the smallest insects.*

USAGE. *Entomology, Zoology, Evolution, Biomechanics, Fluid mechanics*



# *Multiscale simulations of nanodispersed polymer and supramolecular systems*

*Komarov P.V., Khalatur P.G., Malyshev M.D.*
*Lomonosov Moscow State University*

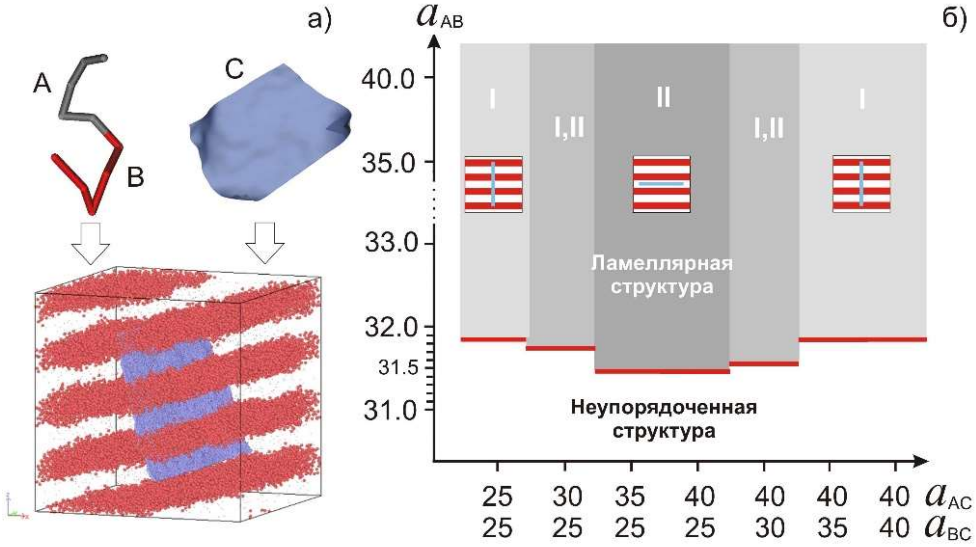
DRIVER. *The study, characterization and virtual design of new functional materials based on the methodology of multiscale simulations.*

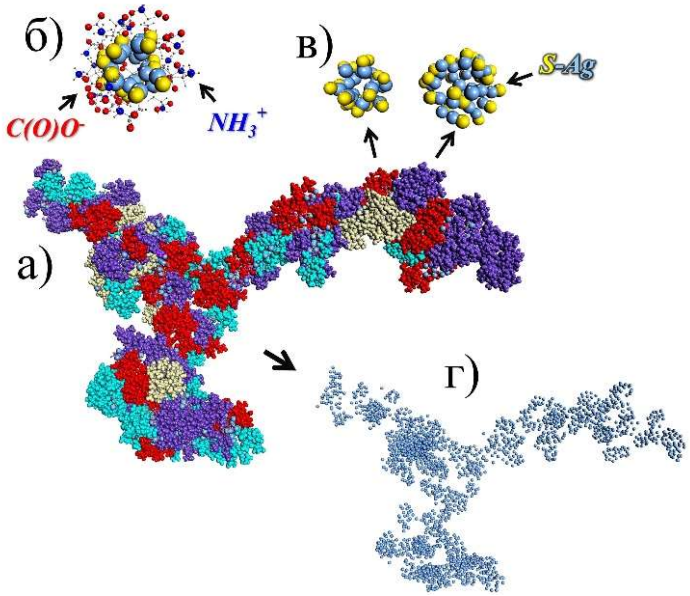
STRATEGY. *This project focuses on the using atomistic and mesoscopic computer simulations such as atomistic molecular dynamics and dissipative particle dynamics. These methods are used to create models of molecular systems and to study their equilibrium / quasi-equilibrium states, which is a necessary step in the context of full multiscale modeling.*

OBJECTIVE. *New functional materials, methods of creation of new materials.*

IMPACT. *Physical chemistry of nanoscale systems, nanomaterials.*

USAGE. *Methods of design and simulation, nanomaterials.*



# *Simulation of branched molecules, microgels and multi-component polymer systems in solvents and melts*

*Gumerov R.A. Lomonosov Moscow State University*

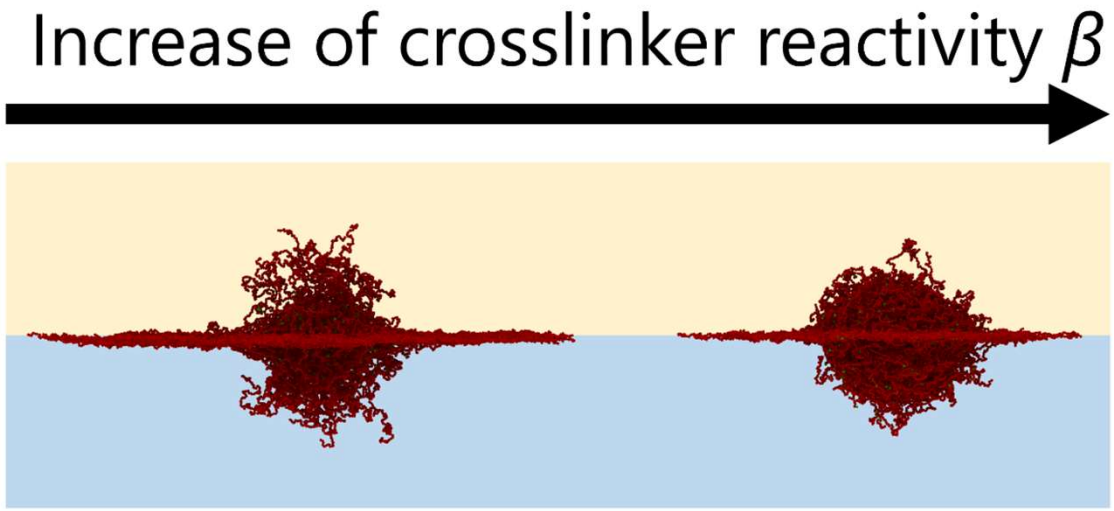
DRIVER. *Study the adsorption of arborescent molecules, microgels on the interfaces and changes of structural characteristics in block copolymer films and nanocomposites during swelling.*

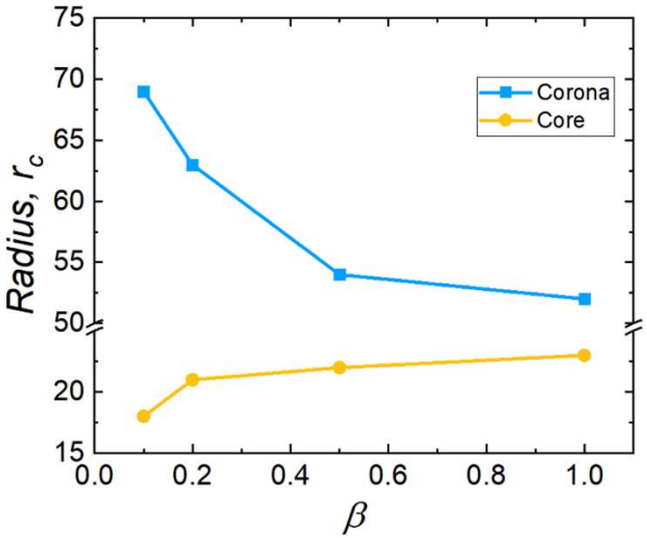
STRATEGY. *Placing a few molecules in a solvent, observe its behavior. The behavior if simulated by the dissipative particle dynamics method.*

OBJECTIVE. *To obtain the stable emulsions and long-range ordered films. In addition, the morphology of molecules and films must correspond to the particular expectances.*

IMPACT. *The microgels and arborescent molecules can cover the bigger area of liquid's interfaces in comparison with linear polymers and surfactants. Swollen thin block copolymer films have inhomogeneous distribution of solvent along the domains.*

USAGE. *High adsorptional properties are used for the obtainment of stable micro- and nanoemultions. Thin films can be used as nanopumps.*



# *Space environment impact on nanostructures and nanomaterials*

*Khlebnikov S.A., Voronina E.N., Chirskaya N.P.*
*MSU SINP*

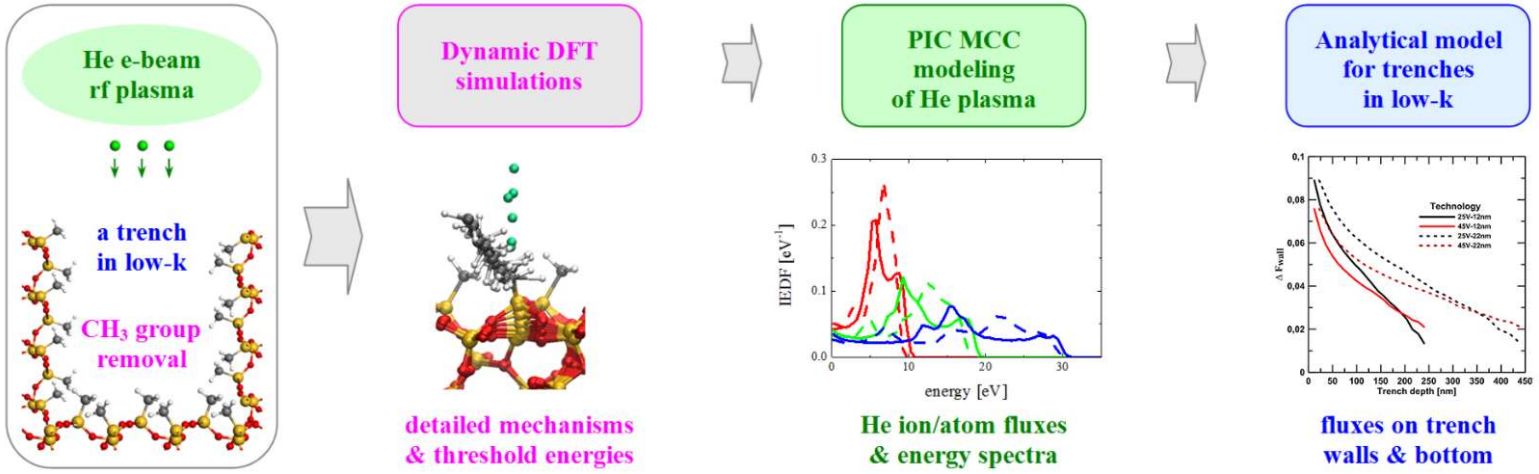
DRIVER. *Development of novel materials with high durability to space environment impact.*

STRATEGY. *Multiscale modeling changes in properties of nanostuctures and nanomaterials under the space environment impact.*

OBJECTIVE. *Study of space environment component impact (particles of cold and hot space plasma, of space radiation, etc.) on nanostructures and nanomaterials. Usage of principles of multiscale simulation for description of space environment components impact on nanostructures and nanomaterials, choice of the set of simulation methods for certain space environment factors as well as their combinations.*

IMPACT. *Forecasting nanostructures and nanomaterials behavior under extreme space conditions. Improvement of spacecraft material durability. Recommendations on development novel materials with high durability to space environment impact.*

USAGE. *Spacecraft design and development, nanotechnology.*



# *Efficient creation of enzymes based on immunoglobulins*

*Golovin A.V. Lomonosov Moscow State University*

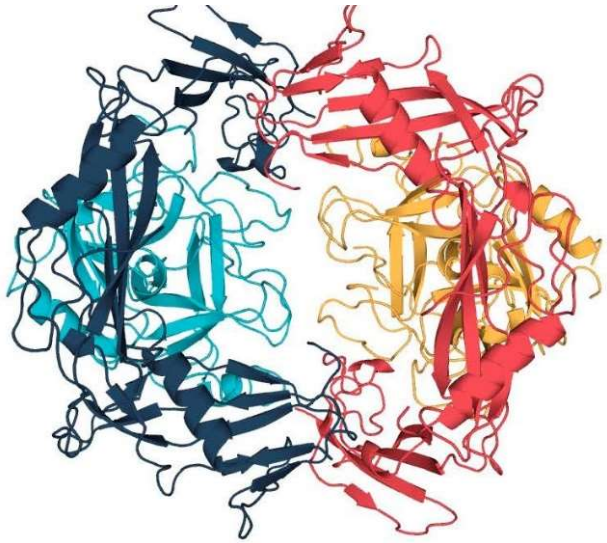
DRIVER. *Creation of new enzymes based on immunoglobulins using molecular modeling techniques.*

STRATEGY. *A complete study of immunoglobulins mutation space in the active site and active site transfer from known enzymes in order to achieve a high level of enzymatic activity comparable with the one of natural enzymes.*

OBJECTIVE. *Creating of computational approaches using methods of molecular mechanics, quantum chemistry, and their hybrids to create new enzymes based on immunoglobulin.*

IMPACT. *Immunoglobulins play an important role in modern biopharmaceutical, giving them catalytic properties will increase their effectiveness in many times.*

USAGE. *Medicine, Pharmacology, Basic Research.*



# *New multicomponent mitotic spindle inhibition strategies to stop tumor cell division*

*Gudimchuk N.B.[1,2], Alexandrova V.V. [1,3], Eltsov I.A.[1,3], Lopanskaya Yu.N.[1,2], Fedorov V.A.[1,2], Anisimov M.N.[1,2]*
*1) Lomonosov Moscow State University, 2) Center for Theoretical Problems of Physical and Chemical Pharmacology RAS, 3) Moscow Institute of Physics and Technology*

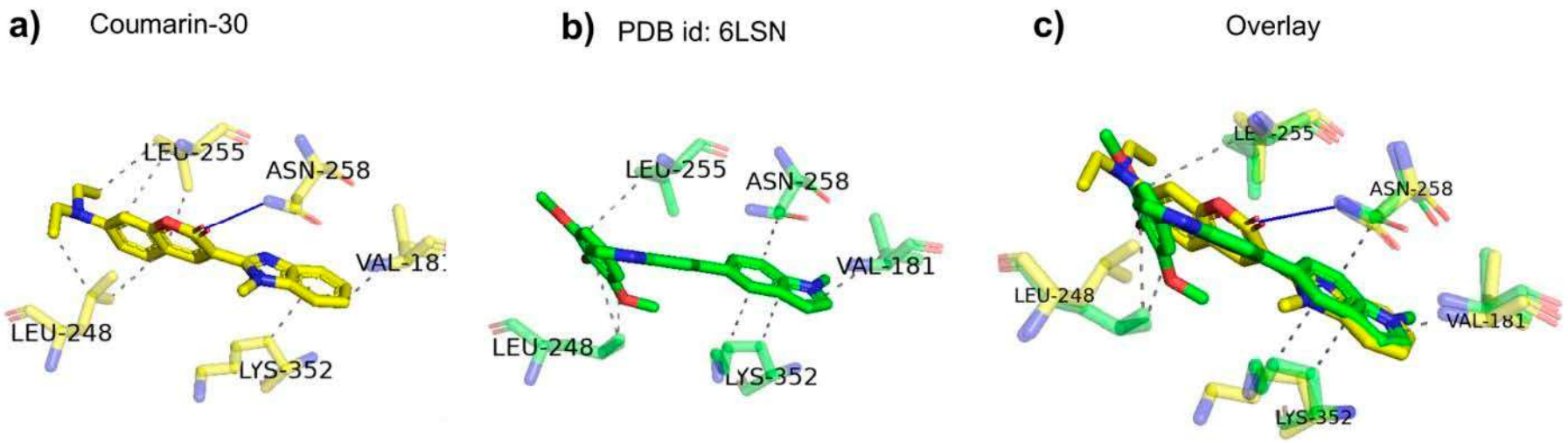
DRIVER. *Development of alternative antitumor drugs that would overcome the body's resistance to currently existing drugs, as well as their side effects.*

STRATEGY. *Finding drugs for alternative targets (microtubule-associated proteins).*

OBJECTIVE. *Development of alternative antitumor drugs and creation of a new method of work in this field.*

IMPACT. *Scientific: the creation of an effective method of drug development will contribute to the field of research on the interaction of proteins and low-molecular-weight compounds.*

USAGE. *1) medicine 2) pharmacology 3) the field of science dealing with the interaction of proteins and low-molecular-weight compounds.*



# *Simulation of complex turbulent flows using unstructured meshes*

*A.P. Duben, I.V. Abalakin, P.A. Bakhlovalov, V.G. Bobkov, A.V. Gorobets, N.S. Zhdanova, T.K. Kozubskaya, P.V. Rodionov Keldysh Institute of Applied Mathematics of RAS*

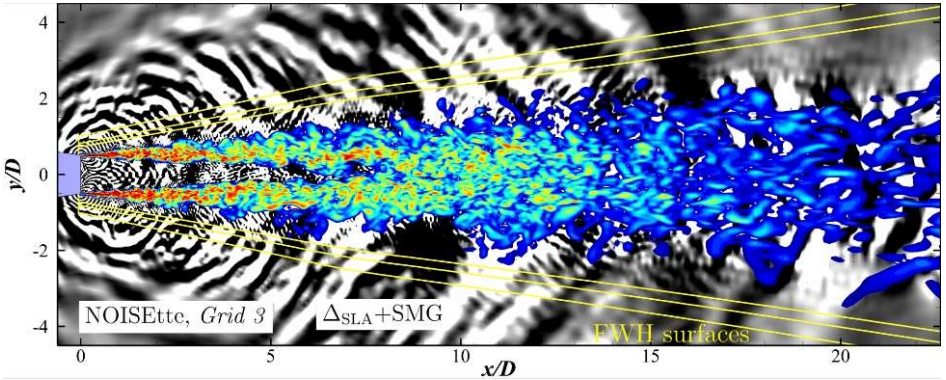
DRIVER. *Numerical investigation of aerodynamic and aeroacoustic charateristics of bluff aircraft elements and complex turbulent flows*.

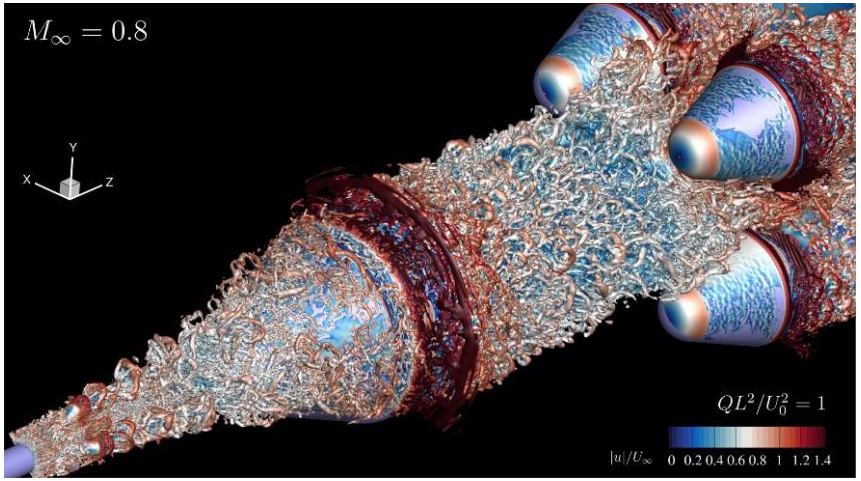
STRATEGY. *Providing computational experiments on different configurations, comparison experimental and numerical data, identification of optimal configurations.*

OBJECTIVE. *Investigation and improvement of aerodynamic and aeroacoustic characteristics of aircraft elements and complex turbulent flows.*

IMPACT. *Improvement of aircraft aerodynamic characteristics; reduce of airframe noise in near and far field.*

USAGE. *Aircraft industry; computational mathematics; computer science.*



# *Theoretical modelling of electronic and magnetic properties of endohedral fullerenes and their derivatives*

*Ioffe I.N., Lukonina N.S., Mazaleva O.N., Sudarkova S.M., Pykhova A.D., Khinevich V.E.*
*Lomonosov Moscow State University*

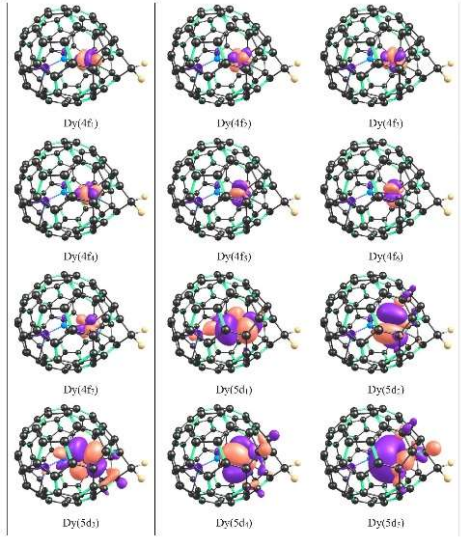
DRIVER. *The present project is aimed at identification of the promising single-molecule magnets based on the endohedral metallofullerenes.*

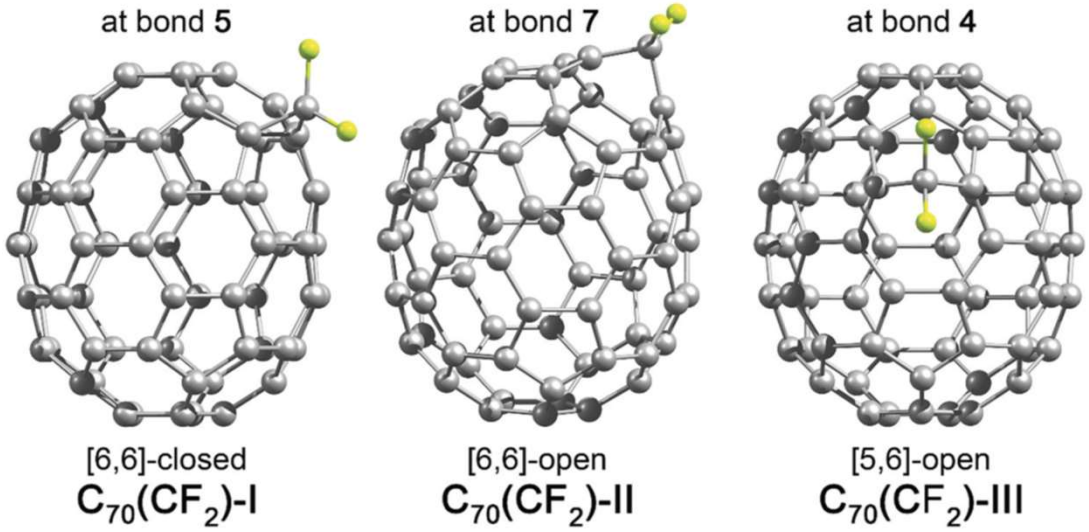
STRATEGY. *All possible orientations of the endohedral atoms within the carbon cage will be considered in order to elucidate the effect of the exohedral addends on their dynamics.*

OBJECTIVE. *Elucidation of the dynamics of the endohedral clusters inside various carbon cages and of its interplay with the exohedral addition patterns.*

IMPACT. *The project will provide better understanding of the fundamental aspects of the electronic and magnetic properties of endohedral fullerenes. It will further help to identify those endohedral fullerenes that may be applicable as novel materials for spintronics, information storage, etc.*

USAGE. *Fundamental and applied material science.*





# *Testing of neural network methods for natural langugage processing*

*Lukashevich N.V. Lomonosov Moscow State University*

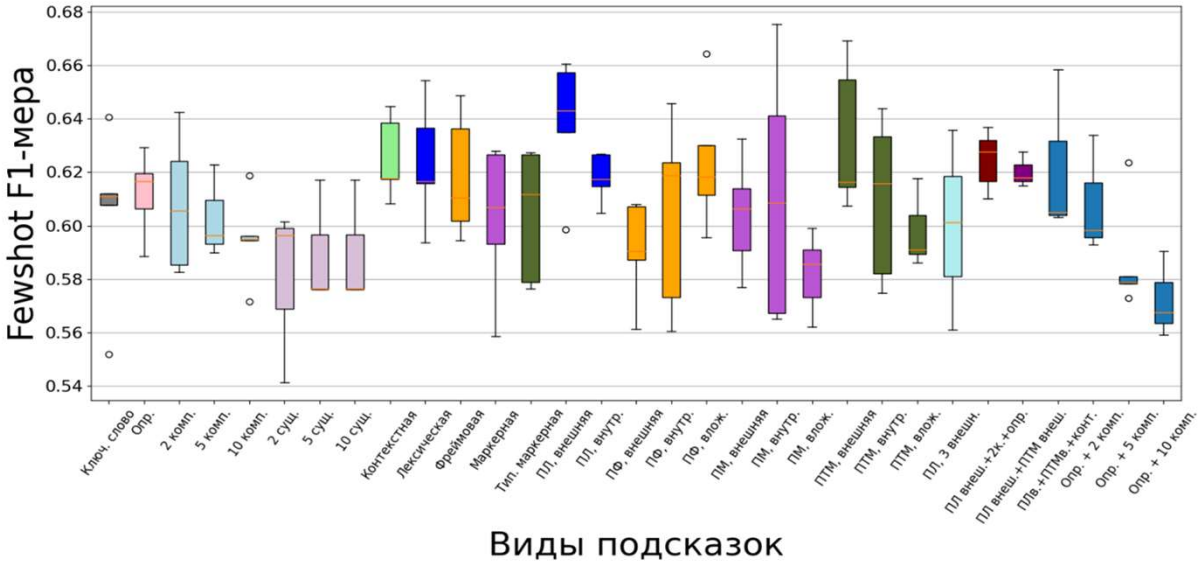
DRIVER. *The study of methods for extracting nested named entities based on two datasets in Russian*.

STRATEGY. *The study of the MRC model (Machine Reading Comprehension model) and questions to this model for extracting named entities.*

OBJECTIVE. *Test deep learning neural network models for extracting information from texts to enrich knowledge graphs.*

IMPACT. *Customization of existing knowledge graphs for specific domains in order to improve the quality of natural language processing in these domains.*

USAGE. *Natural language processing systems, information-analytical systems.*



# *Molecular modeling of biochemical and biophysical processes in cholinesterases and development of new drugs interacting with them*

*Lushchekina S.V., Novichkova D.A., Cabbage D.P., Kots E.D., Boyko K.M. Institute of Biochemical Physics. N.M. Emanuel RAS, Lomonosov Moscow State University*

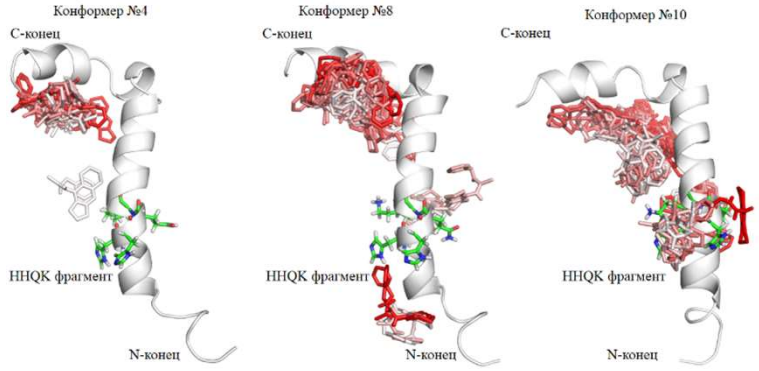
DRIVER. *Cholinesterases play crucial role in human physiology. Detailed understanding of their biochemical and biophysical properties is crucial for better understanding of human body metabolism. Cholinesterases are involved in development of several serious conditions (Alzheimer disease, myasthenia) and as such are targets for drug treatment of these diseases. Also there are targets of poisonous compounds and better understanding of their functioning mechanism is necessary for development of protective measures.*

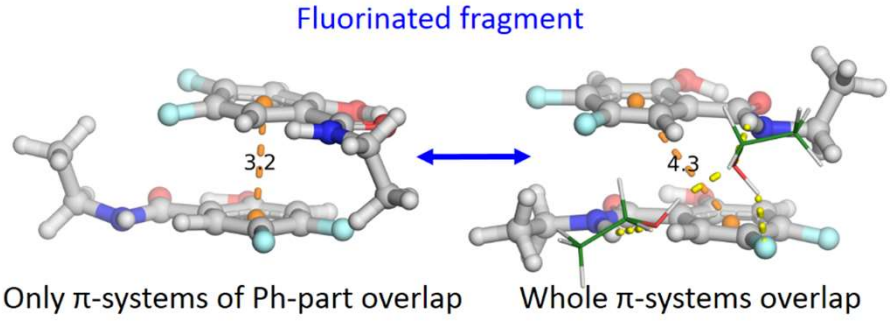
STRATEGY. *Molecular modeling of dynamics and conformational changes in the enzymes protein molecules, study of energy profiles of their interactions with inhibitors and substrates by means of molecular docking, molecular dynamics, quantum mechanics, combined quantum mechanics and molecular mechanics, combined quantum mechanics and molecular dynamics methods.*

OBJECTIVE. *New information about processes in biological systems and their interactions with different chemical compounds – poisons and drugs. Development of new drugs and protective agents.*

IMPACT. *Extension of basic knowledge for biochemical and biomedical research. Development and implementation of new drugs.*

USAGE. *Pharmacology and medicine. Anti-Alzheimer disease and anti-myasthenic drugs, protective agents for aviation (passengers, crew, land personnel) against aerotoxic syndrome.*



# *Structure and dynamic properties of polymer nanocomposites*

*Lyulin S.V., Gurtovenko A.A., Larin S.V., Fal'kovich S.G., Nazarichev V.M., Volgin I.V., Glova A.D. Institute Of Macromolecular Compounds RAS*

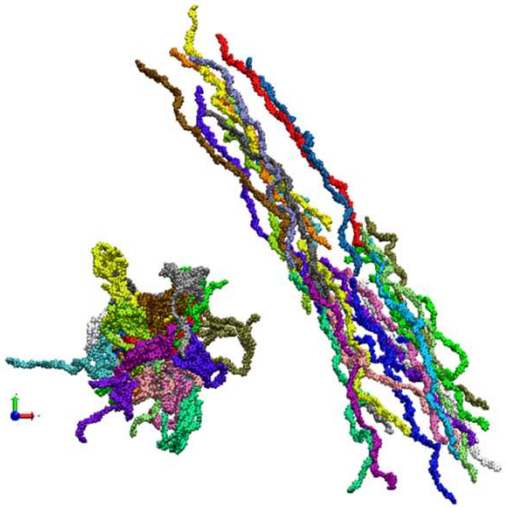
DRIVER. *Study and prediction of structure and dynamic properties of polymer nanocomposites.*

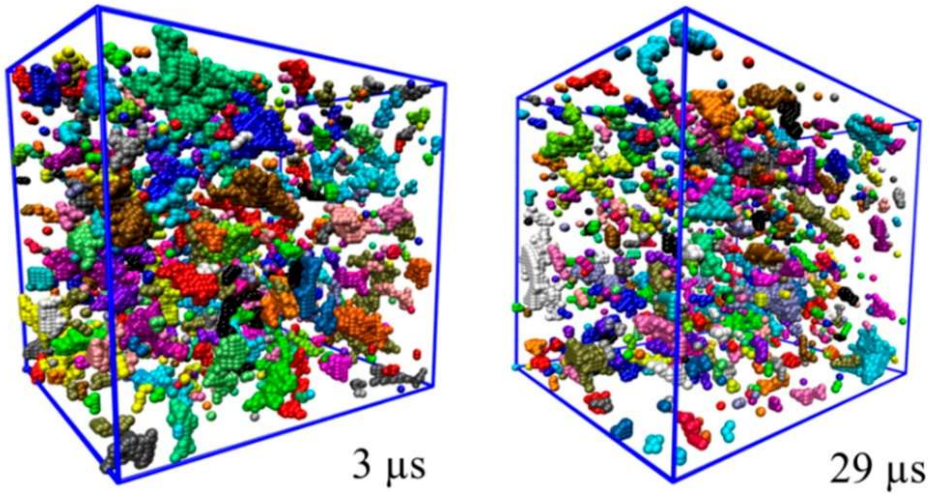
STRATEGY. *Development of nanocomposite models, further computer simulation of their structure and dynamic properties on the base of quantum mechanics and full-atomic molecular dynamics.*

OBJECTIVE. *Virtual design and testing of advanced polymer nanocomposites, including development of nanocomposite models, further computer simulation of their structure and dynamic properties, study of reinforcement mechanisms.*

IMPACT. *Methodology of calculation and prediction of polymer nanocomposites properties.*

USAGE. *Development of new polymer nanocomposites.*





# *Fundamental studies of combustion and detonation processes in relation to the development of the foundations of energy technologies*

*Levin V.A., Manuilovich I.S., Zhuravskaya T.A., Sutyrin O.G.*
*Lomonosov Moscow State University*

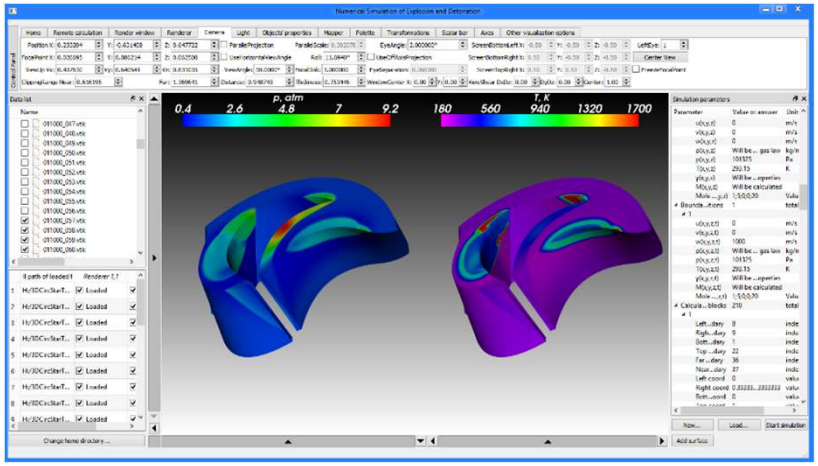
DRIVER. *Study of the detonation waves in inhomogeneous mixtures, including investigation of detonation dynamics and its structure in channels of complicated geometry and in combustion chambers*.

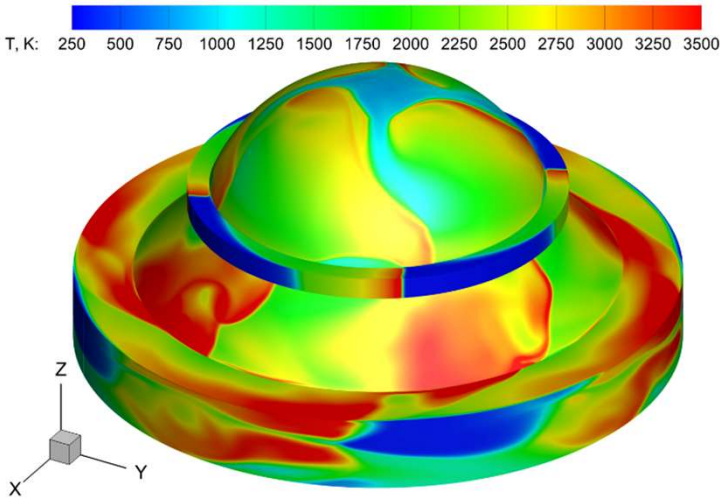
STRATEGY. *Fast parallelized simulation of reactive multicomponent flows in domains of complicated geometry.*

OBJECTIVE. *Create original software package intended to simulate 1D, 2D and 3D multicomponent reactive flows.*

IMPACT. *Fast and easy simulation of complicated 1D, 2D and 3D multicomponent reactive flows with combustion and detonation.*

USAGE. *Development of new types of engines, solving problems related to explosion safety at various facilities.*



# *Supercomputer modeling of biomolecular systems based on quantum chemical and molecular dynamic methods: from enzymatic catalysis to optogenetics*

*Nemukhin A.V. Lomonosov Moscow State University*

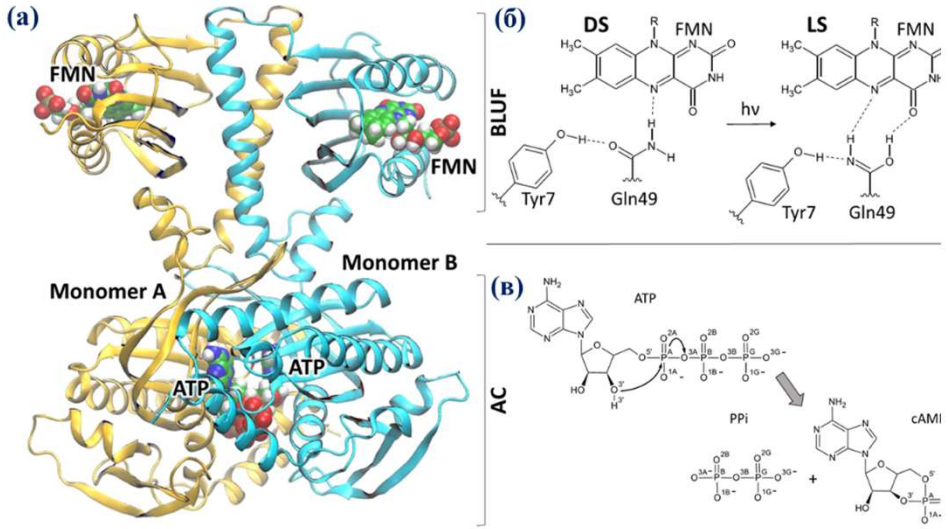
DRIVER. *Development and application of methods for modeling the structure and properties of biomolecular systems.*

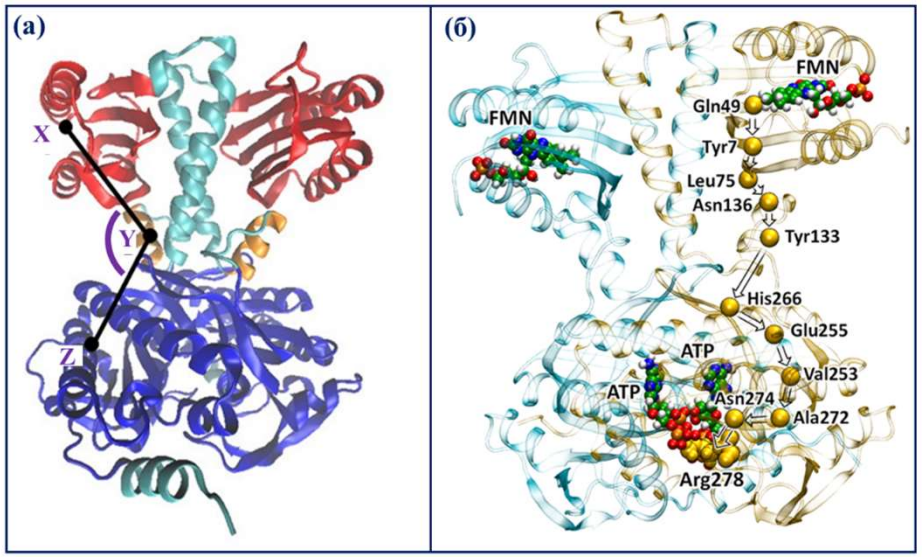
STRATEGY. *The practical implementation of the combined method of quantum mechanics / molecular mechanics (QM / MM) allows us to reach a qualitatively new level in computer modeling of the properties of biomolecular systems.*

OBJECTIVE. *Solve specific problems of modeling biomolecules using the methods of quantum chemistry, KM / MM and MD, focused on supercomputer calculations.*

IMPACT. *will study the molecular mechanisms of action of biomolecular systems necessary to develop ways to control the processes in these systems.*

USAGE. *Biomedicine, biotechnology.*



# *The fundamental research effort on post-genom investigations and technologies*

*Kutov D.K.,Ilyin I.S., Tashchilova A.S., Sulimov V.B., Sulimov A.V.*
*Lomonosov Moscow State University*

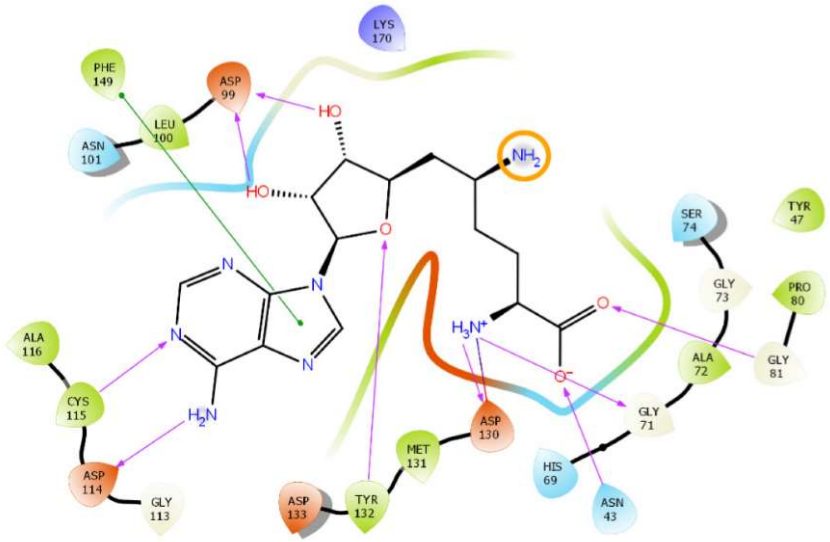
DRIVER. *Design of new inhibitors of urokinase based on molecular modeling.*

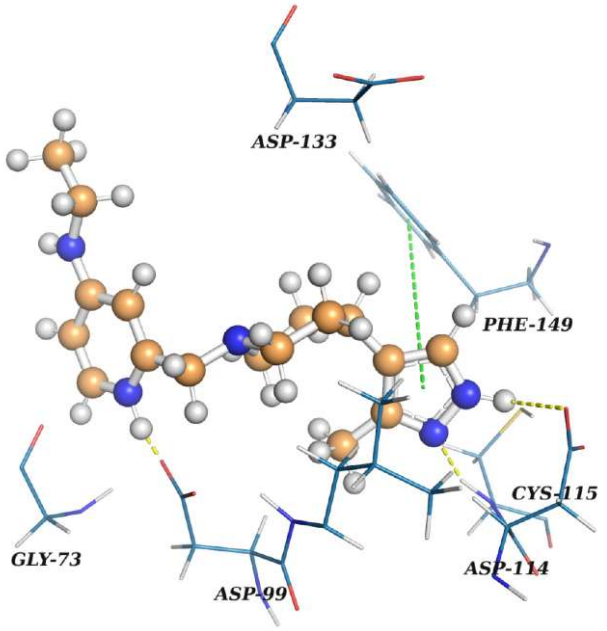
STRATEGY. *Virtual screening of databases of ligands by means of docking, quantum chemistry and molecular dynamics methods for estimation of protein-ligand binding energies.*

OBJECTIVE. *Create original methods based on molecular modeling tools with high ability to predict efficient inhibitors for a urokinase. Higher accuracy of the protein-ligand binding energy calculations leads to the better prediction of new inhibitors. Effective virtual screening of millions of molecules at supercomputers, effective parallel implementations of docking and quantum chemistry methods lead to acceleration of calculations.*

IMPACT. *Efficient, quick and cheap rational development of new drugs and new inhibitors of urokinase.*

USAGE. *Methods lead to new inhibitors development for a given target-protein, in particular urokinase, and new inhibitors of urokinase underlie new drugs.*



# *Modeling of chromatin dynamics*

*Armeev G.A., Kosarim N.A., Gorkovets T.K., Novoseletsky V.N., Pospelova Yu., Andreeva E., Knyazeva A., Shaytan A.K. Lomonosov Moscow State University*

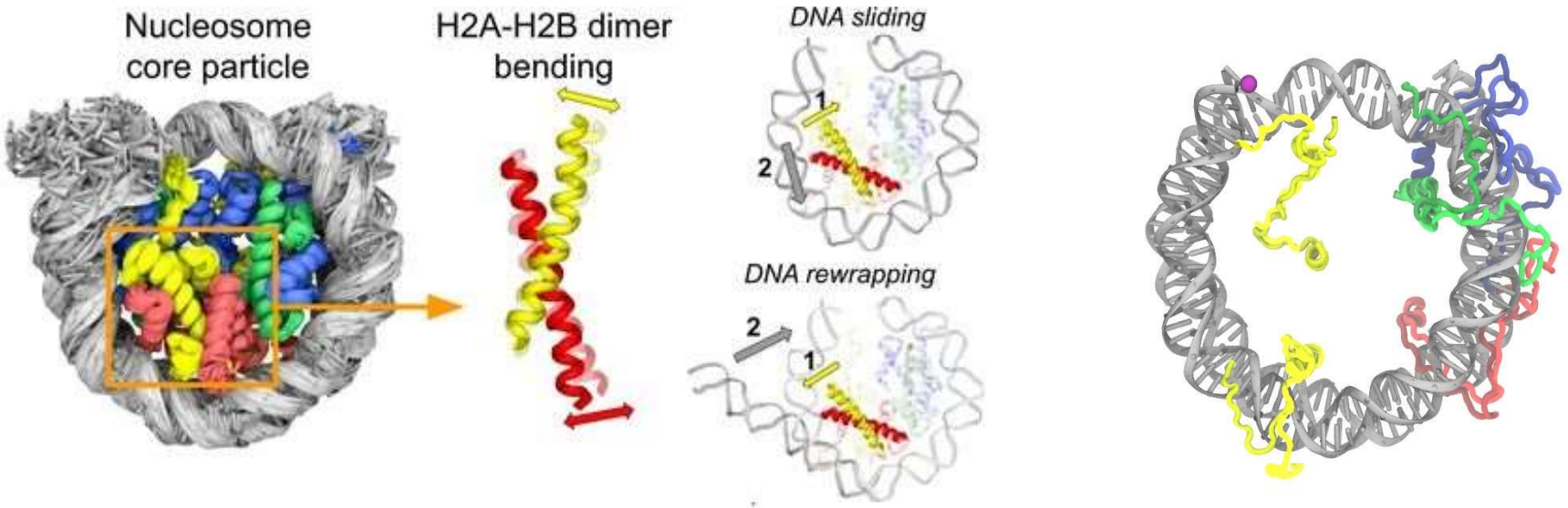
DRIVER. *Understanding the organization and dynamics of protein-protein and protein-DNA interactions in eukaryotic cell nucleus.*

STRATEGY. *Molecular modeling (molecular dynamics, protein-protein docking, Poisson-Boltzman calculations) of protein and DNA interactions in chromatin.*

OBJECTIVE. *Perform complex investigation of key interactions between proteins and DNA at the level of elementary units of chromatin - nucleosomes. Estimate the influence of different interactions and modifications on nucleosome stability and conformation, and as as a result, on the regulation of living organisms functioning.*

IMPACT. *Elucidation of new targets for drug design.*

USAGE. *Medicine, biology, biotechnology, epigenetics.*



# *Molecular modelling of biocatalytic processes*

*Baldin S.M., Kopylov K.E., Drobot V.V., Bochkova A.A., Kirilin E.M., Podshivalov D.D., Shvyadas V.K. Lomonosov Moscow State University*

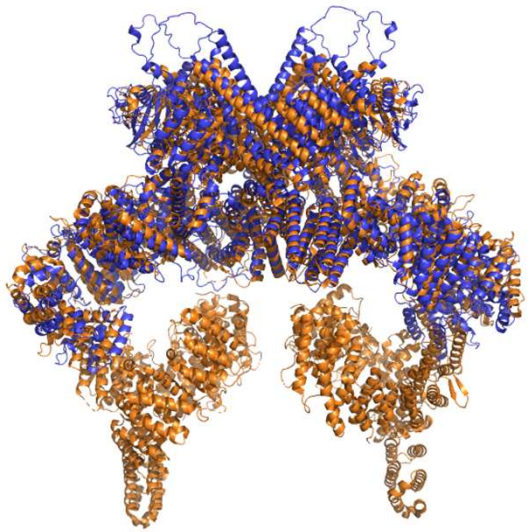
DRIVER. *Obtaining detailed information about the chemical mechanism of enzymes by using molecular modeling*.

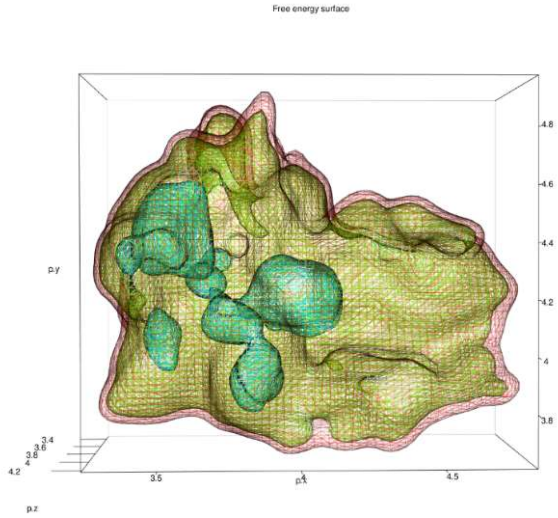
STRATEGY. *Original integrated methodology of molecular modelling including molecular docking, classical molecular dynamics, metadynamics and combined quantum mechanics and molecular mechanics approach (QM/MM).*

OBJECTIVE. *The study of the mechanism of action of enzymes, the possibility of modifying their physico-chemical properties and inhibition.*

IMPACT. *Improving the understanding of the chemical mechanism of enzyme functioning, the dynamics of the catalytic process in active centers and its relation to enzyme function.*

USAGE. *Biochemistry, bioengineering, enzymology, physical chemistry.*

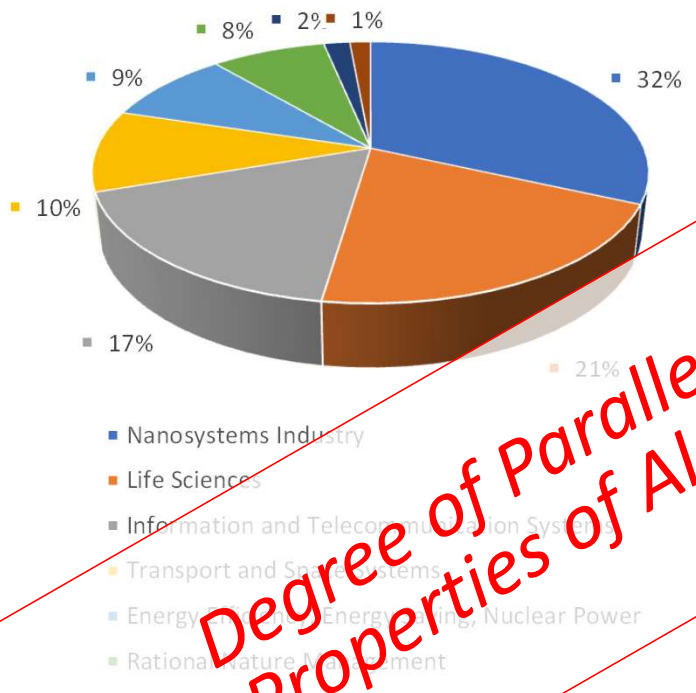




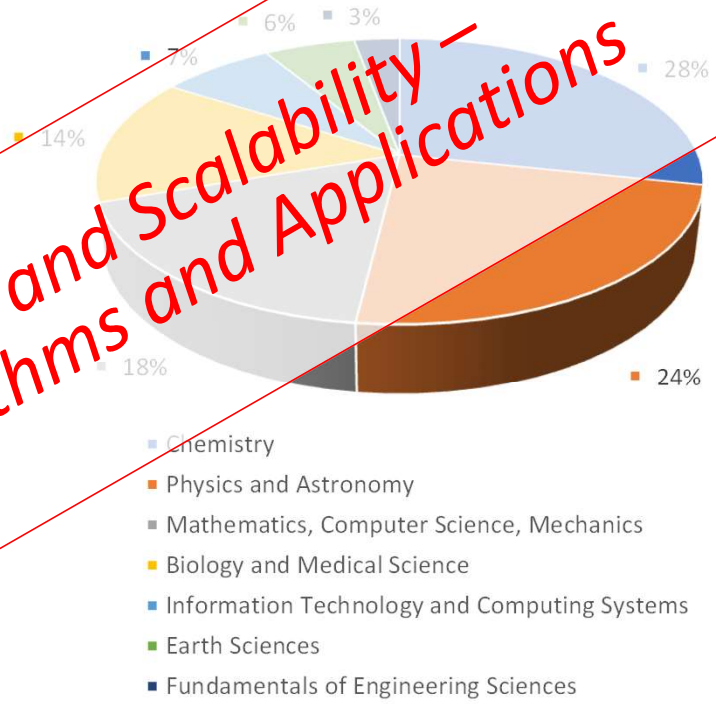
# Supercomputer Center of Moscow University

*(brief statistics for 2022)*

Number of projects according to the priorities of the scientific and technological development strategy




Distribution of projects by branches of science




Degree of Parallelism and Scalability –
Key Properties of Algorithms and Applications

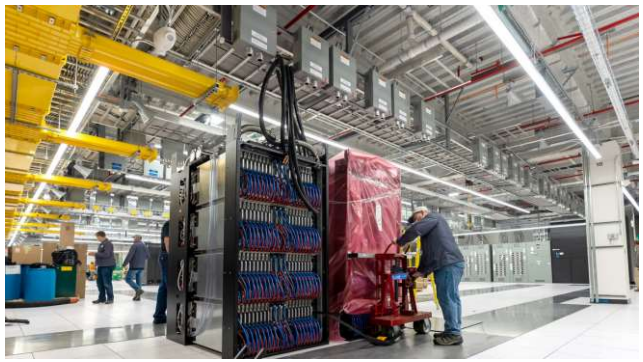
Users – 575
Projects – 200
RSF and RFBR grants - more than 80
Projects of the Ministry of Education and Science - more than 20
Publications – 240
Dissertations (PhD&DrSci) - 16

Departments of Moscow State University - 72%
External organizations - 28%

RSF-OI-2022 projects - 8
(it is maximum among objects of RSF scientific infrastructure)

# *The Frontier Supercomputer*

*(#1 Top500 in 2022-2023 г.)*

*9 408 compute nodes,*
*Each node:*
*1 x CPUs (AMD "Trento", 64 cores, 2GHz)*
*4 x GPU (AMD Radeon MI250X)*
*8 699 904 cores*

*Performance:*
*Peak (theory): 1.68 Eflop/s*
*Linpack test: 1.19 Eflop/s (84%)*

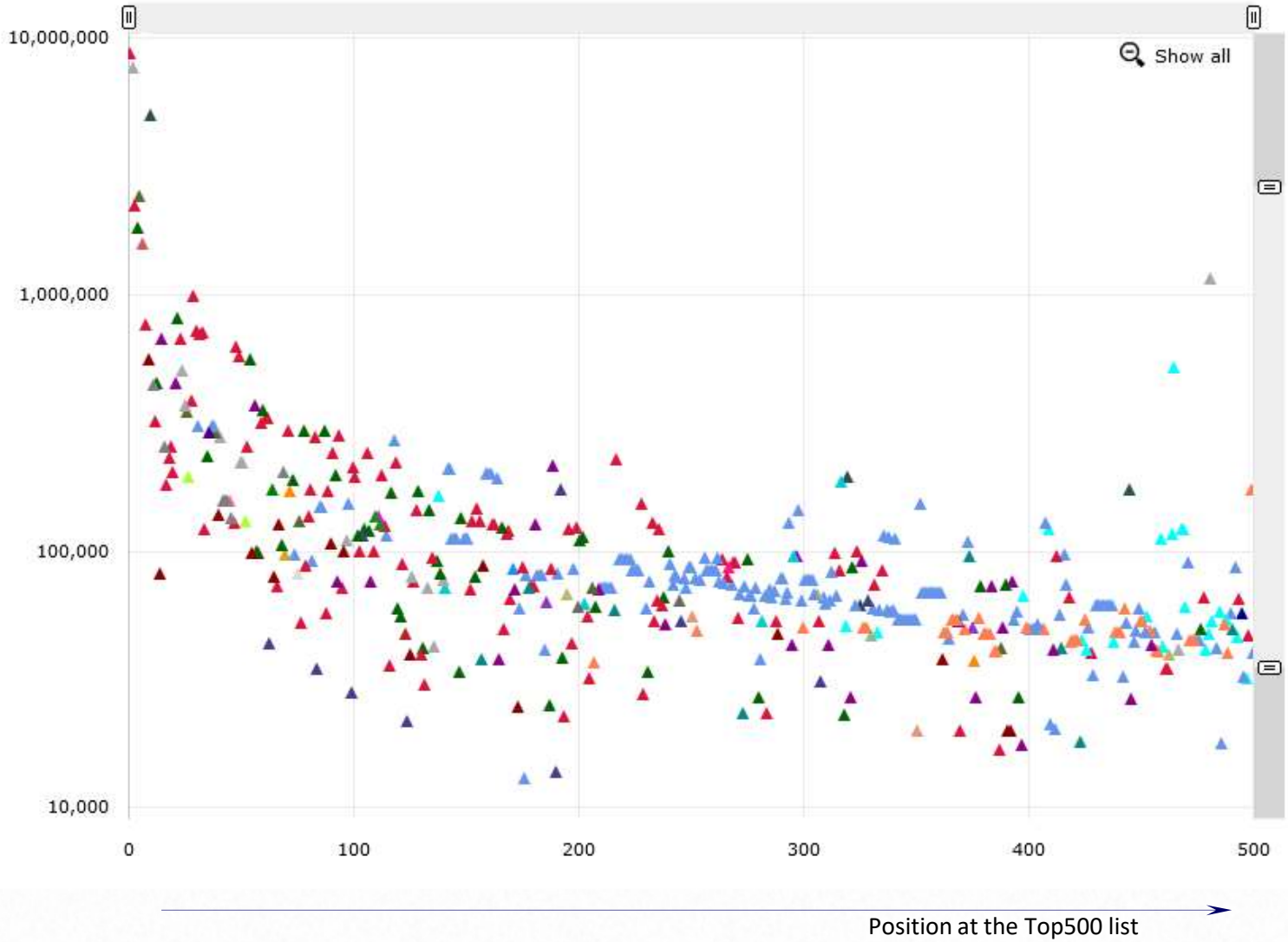
*Main memory = 9.2 PBytes*
*HDD = 716 PB (+37 PB on nodes)*

*22,7 MW (total 29 MW) – 52.2 Gflops/Watt*
*(1 cabinet – 62.68 Gflops/Watt)*

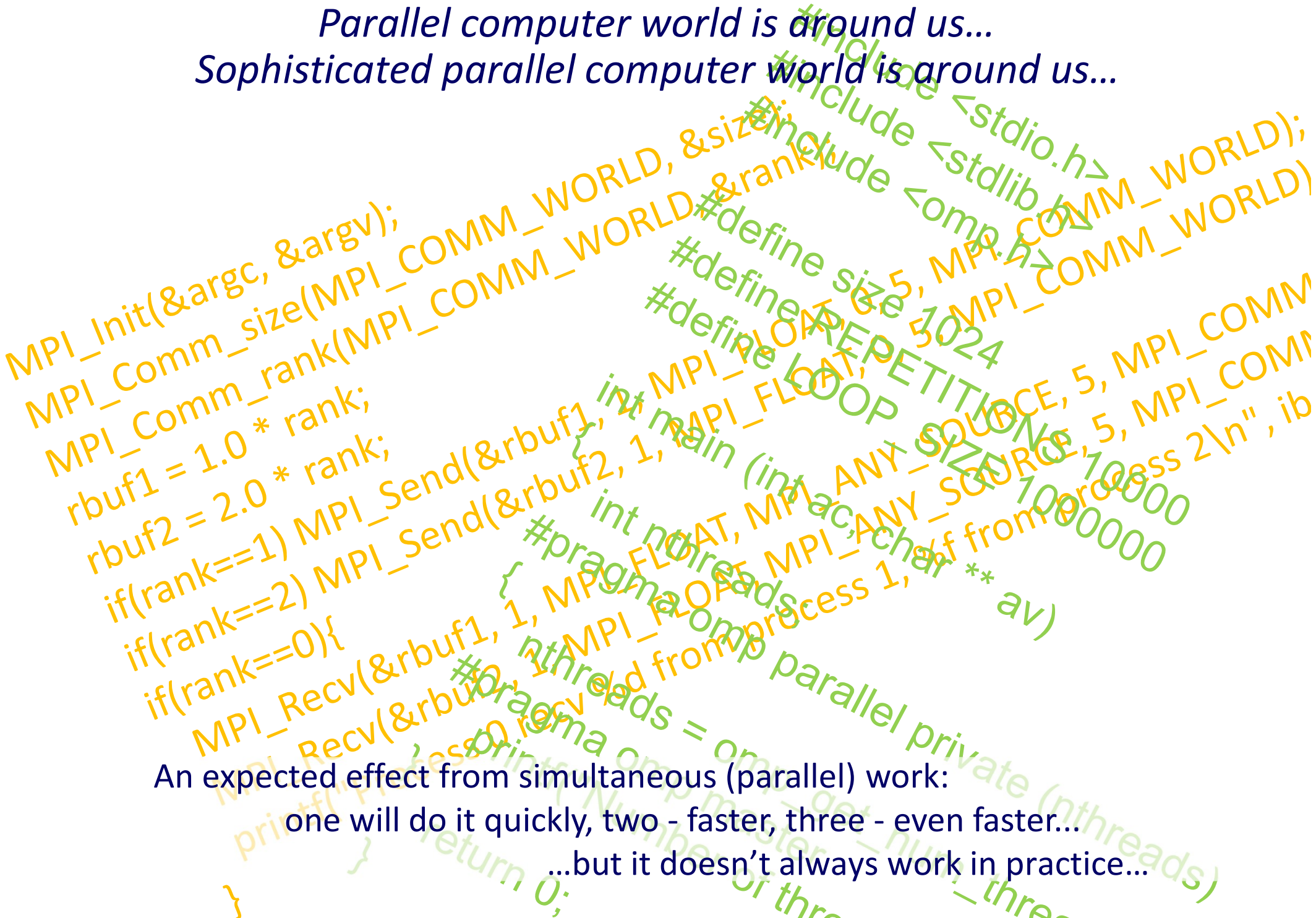
#1 – Top500
#1 – Green500
#1 – HPL-AI

# *Degree of Parallelism: Number of Cores in Top500 Systems*

*(http://top500.org, June, 2023)*



*Parallel world is around us…*
*Parallel computer world is around us…*
*Sophisticated parallel computer world is around us…*

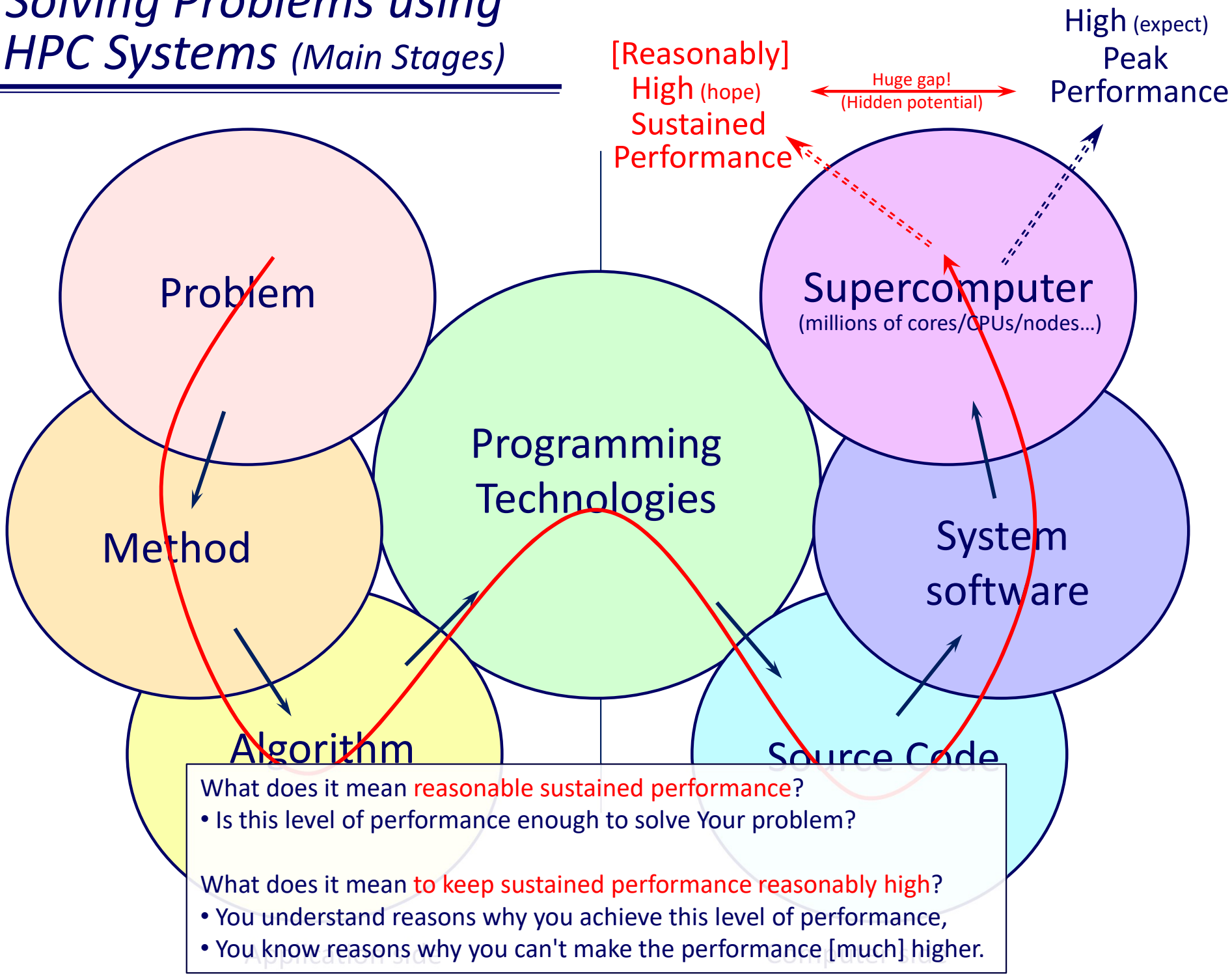



An expected effect from simultaneous (parallel) work:

one will do it quickly, two - faster, three - even faster…

…but it doesn't always work in practice…

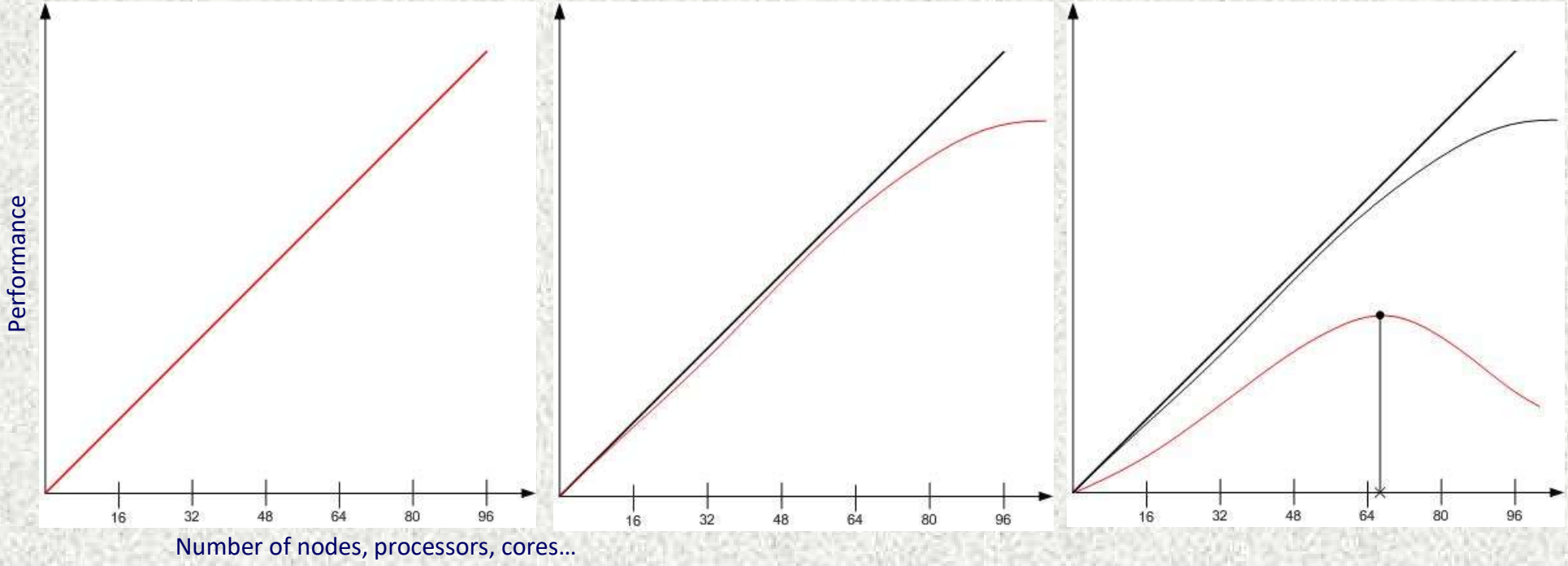
# *Solving Problems using HPC Systems (Main Stages)*




What does it mean reasonable sustained performance?
- Is this level of performance enough to solve Your problem?

What does it mean to keep sustained performance reasonably high?
- You understand reasons why you achieve this level of performance,
- You know reasons why you can't make the performance [much] higher.

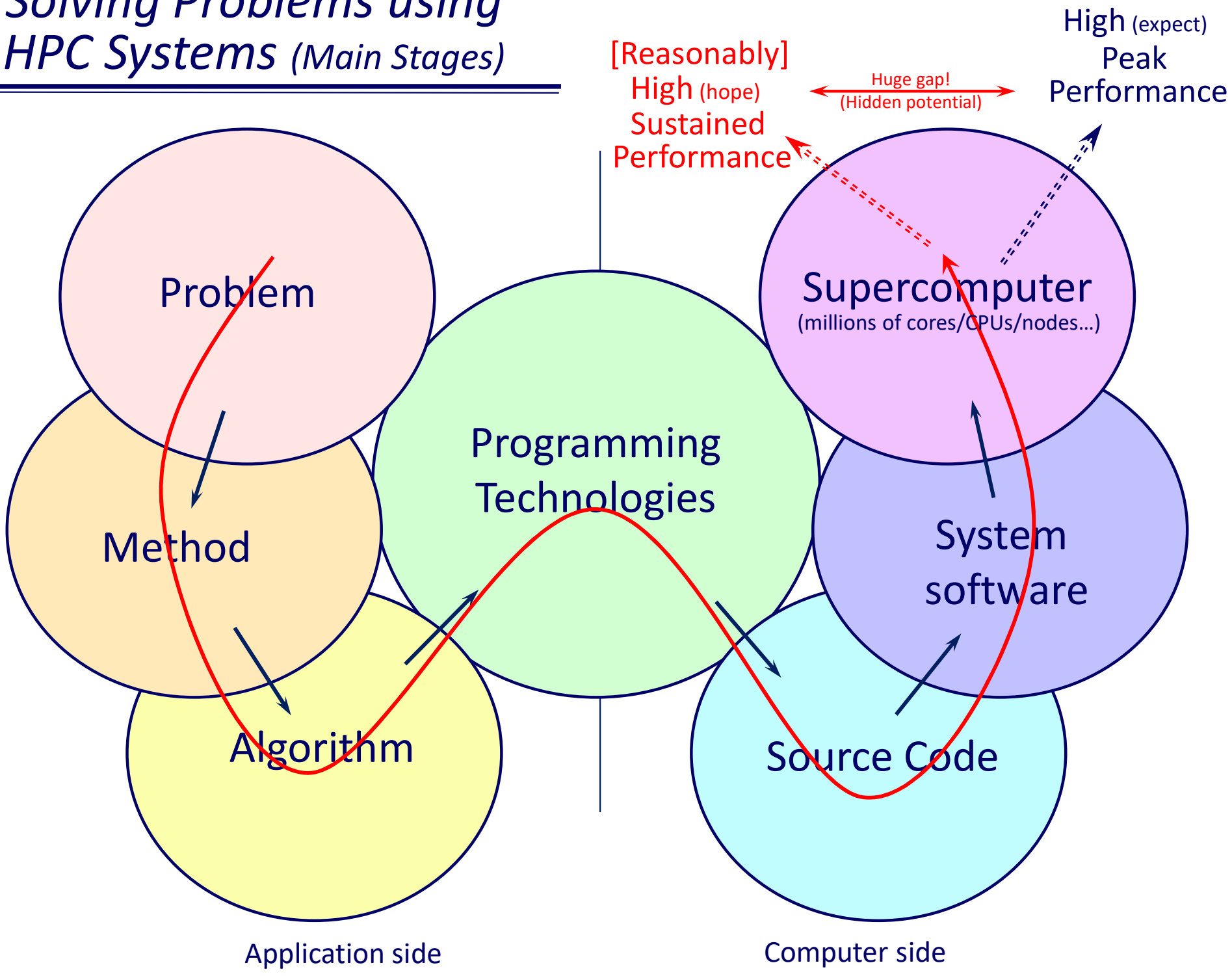
# *Scalability of parallel applications and parallel computing systems*
*(all stages matter!)*



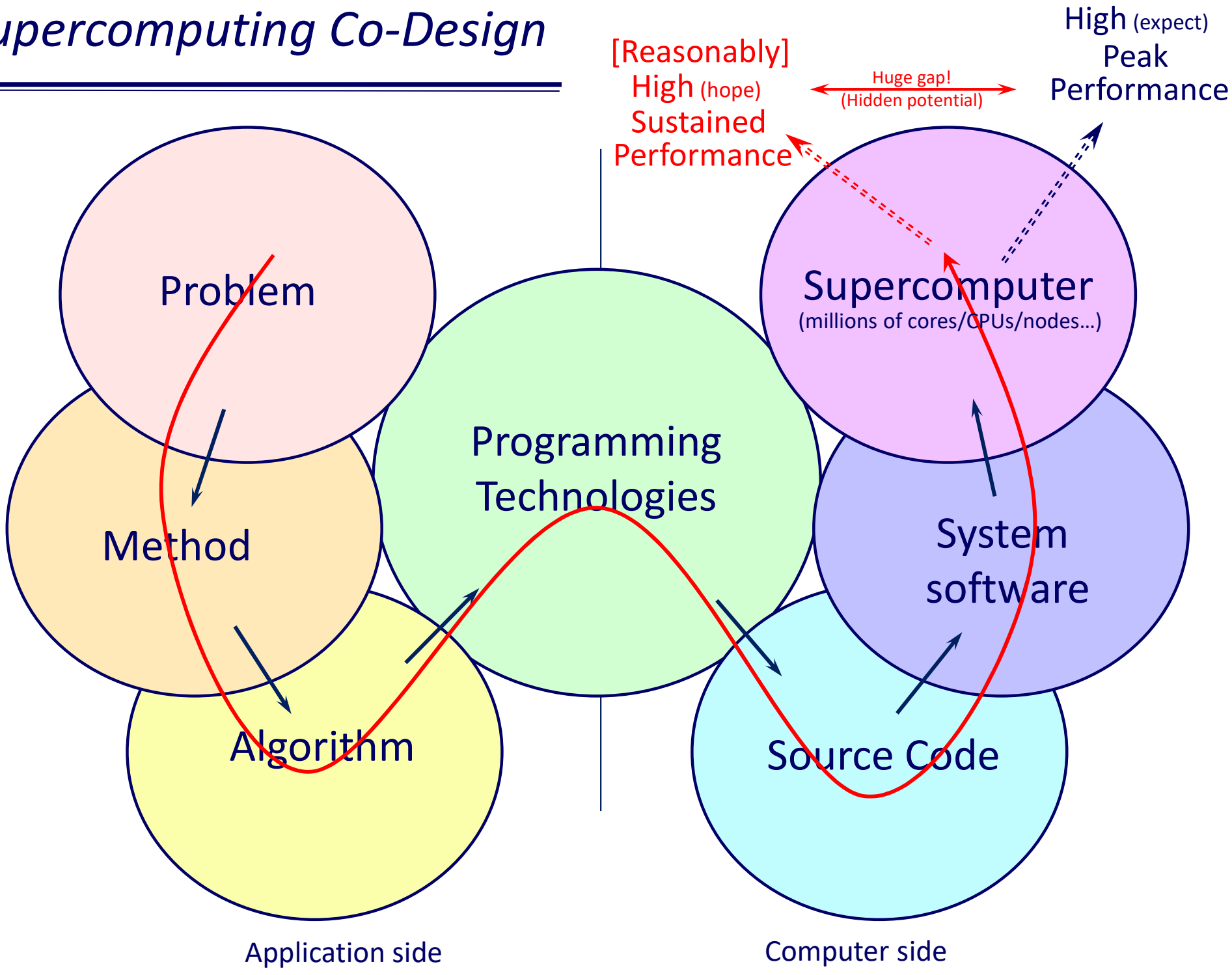

*Possible Causes of Scalability Degradation:*

- *Multiple spawn/destruction of threads.*
- *Unnecessary barriers to thread synchronization.*
- *Load imbalance.*
- *Multiple conflicts when data sharing.*
- *Limited resource of parallelism.*
- *Noise of software environments.*
- *Weak overlapping of computations and communications.*
- *Unnecessarily fine granularity in data transfer.*
- *No binding of threads to cores.*
- *and many more…*

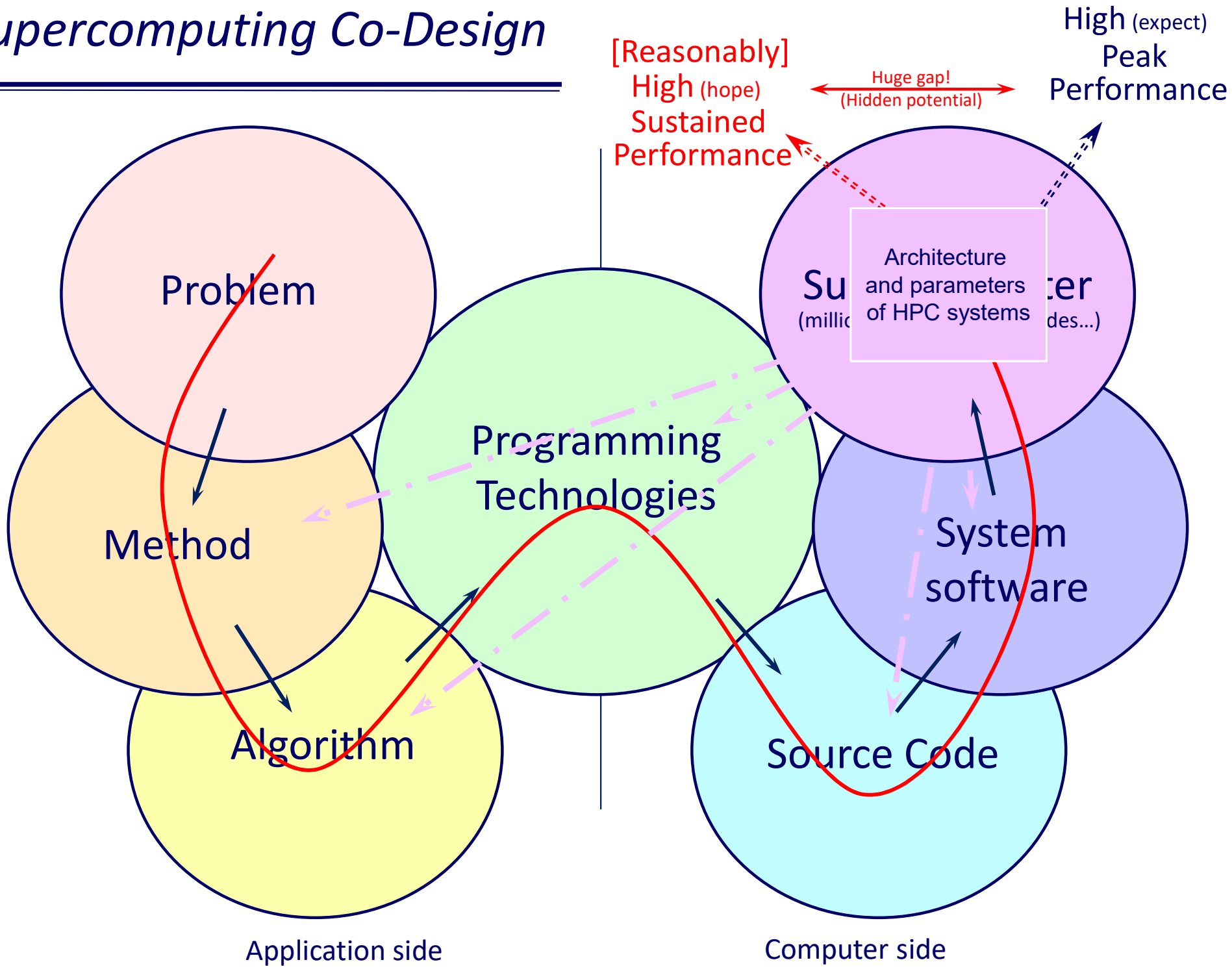
# *Solving Problems using HPC Systems (Main Stages)*



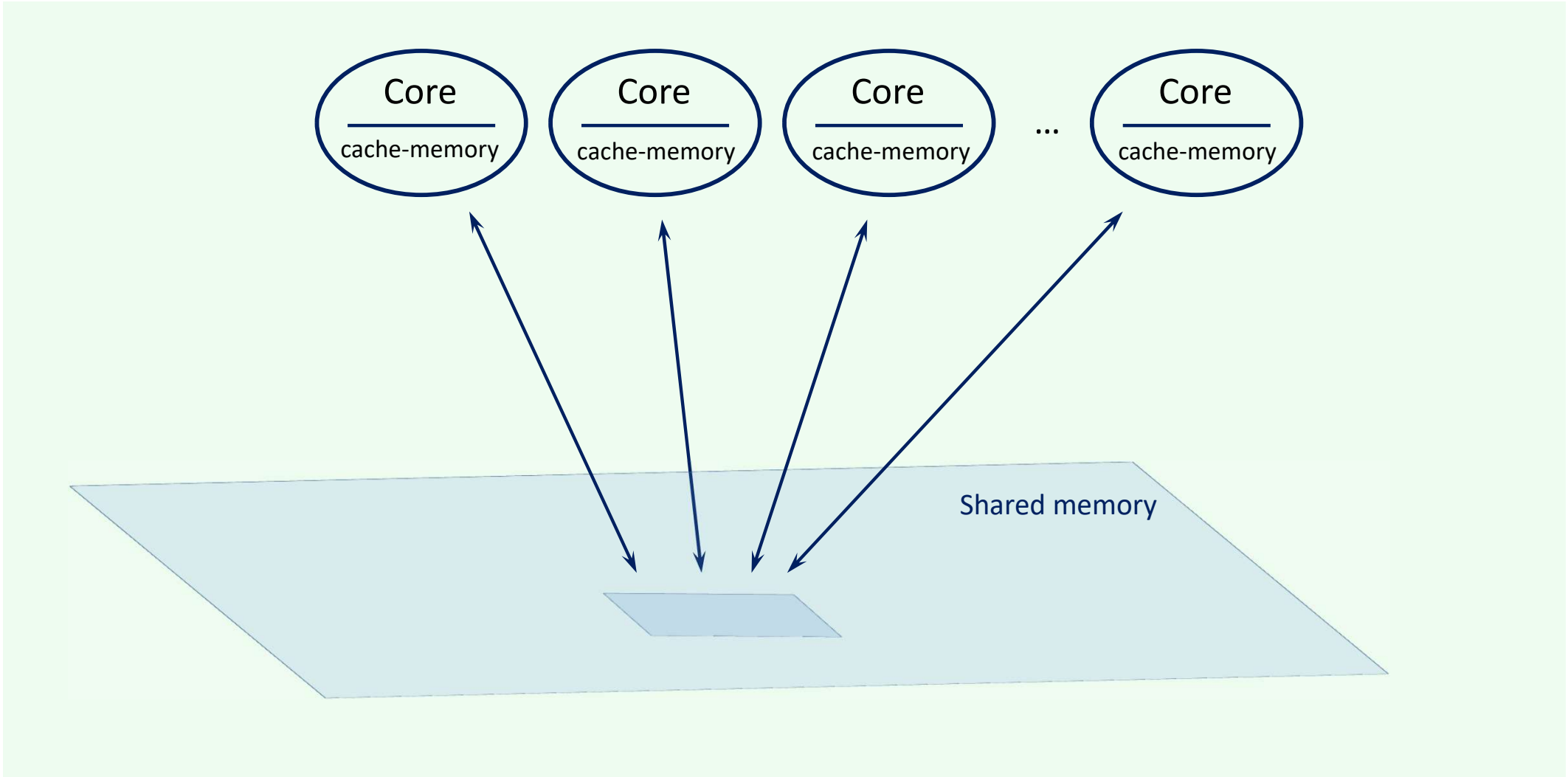
# *Supercomputing Co-Design*



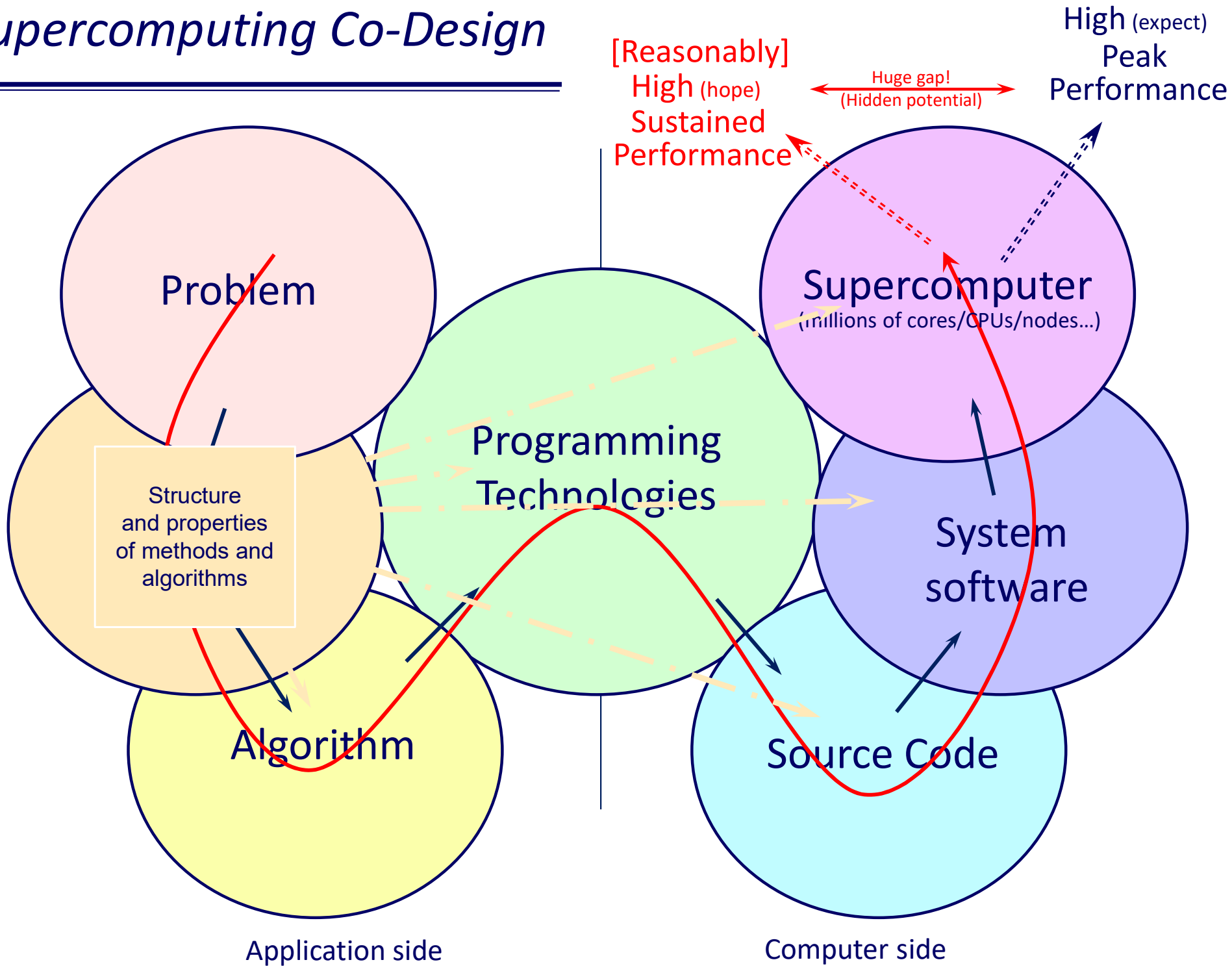
# *Supercomputing Co-Design*



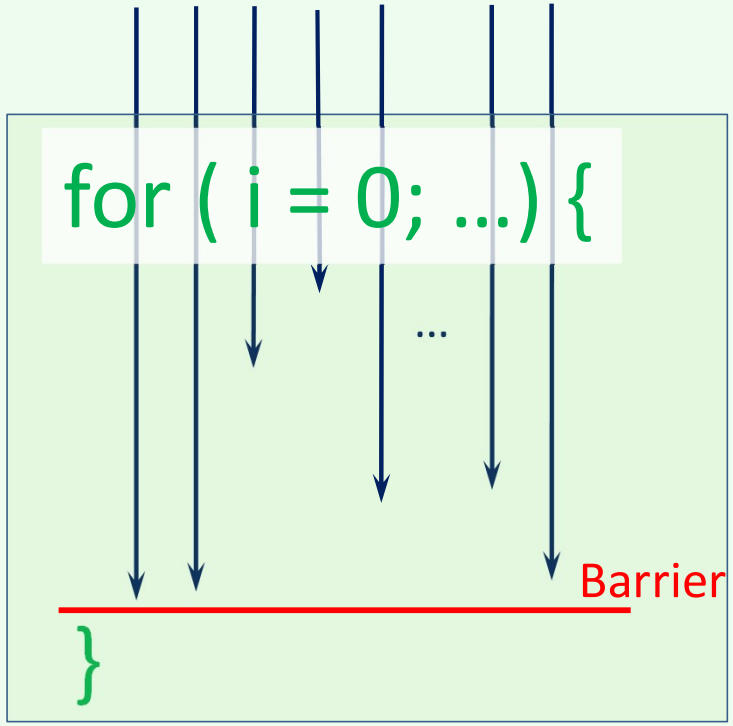
# *Accessing the same shared memory data*
## *(efficiency of ccNUMA protocols)*



# *Supercomputing Co-Design*

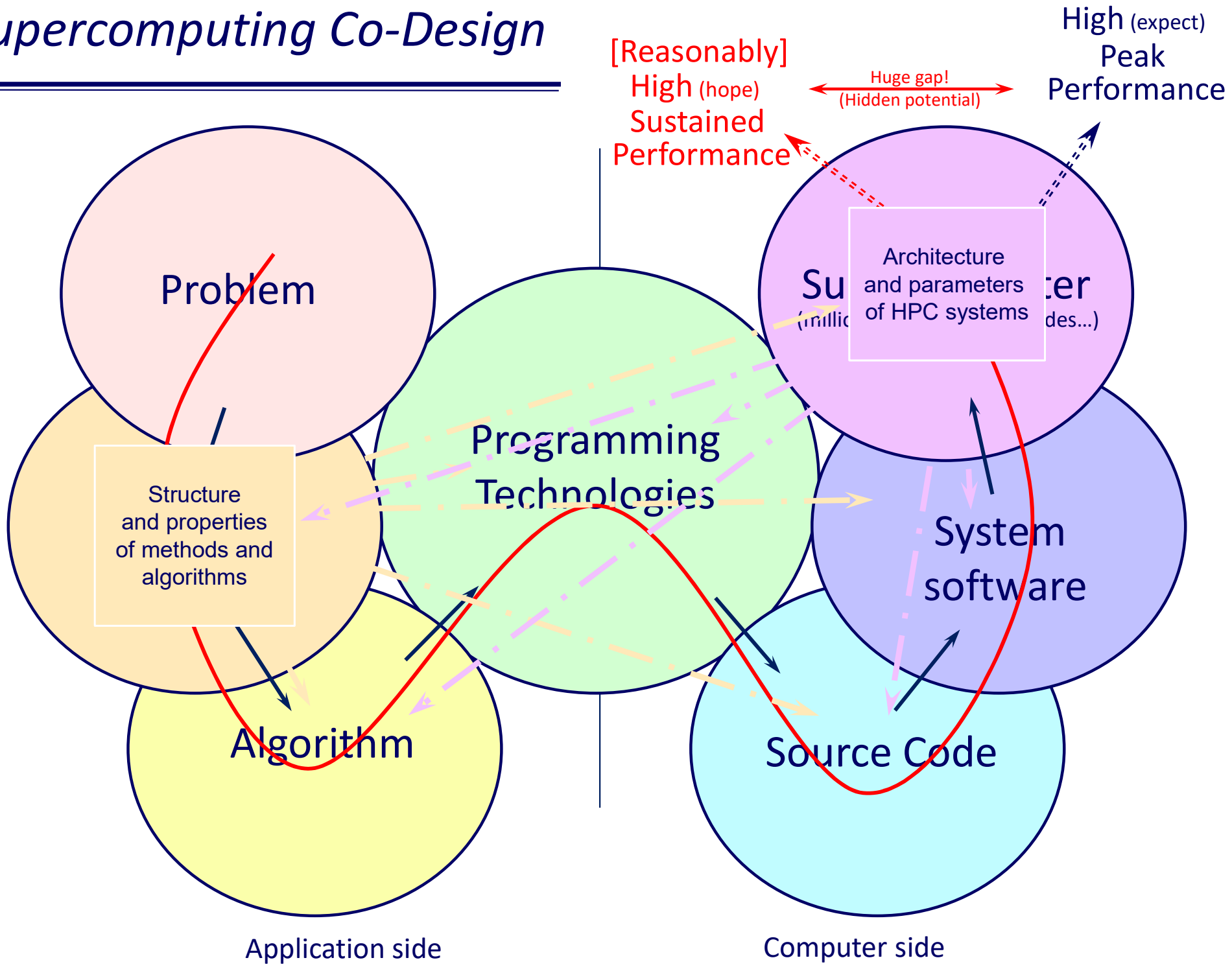


# *Imbalance of Parallel Processing and Processes*

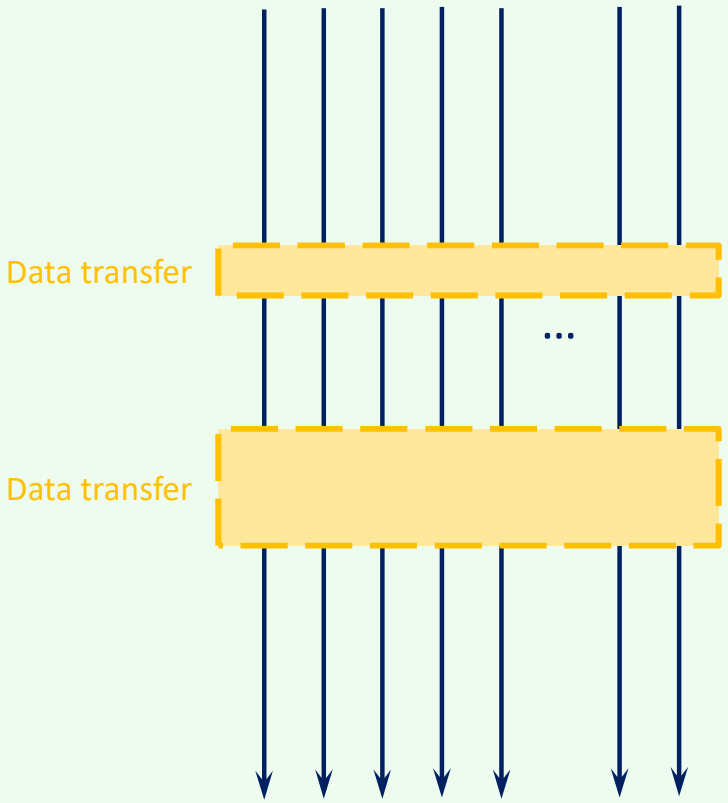



Important: in practice, there are a wide variety of reasons that cause imbalance, from properties of algorithms and data structures to features of the software stack

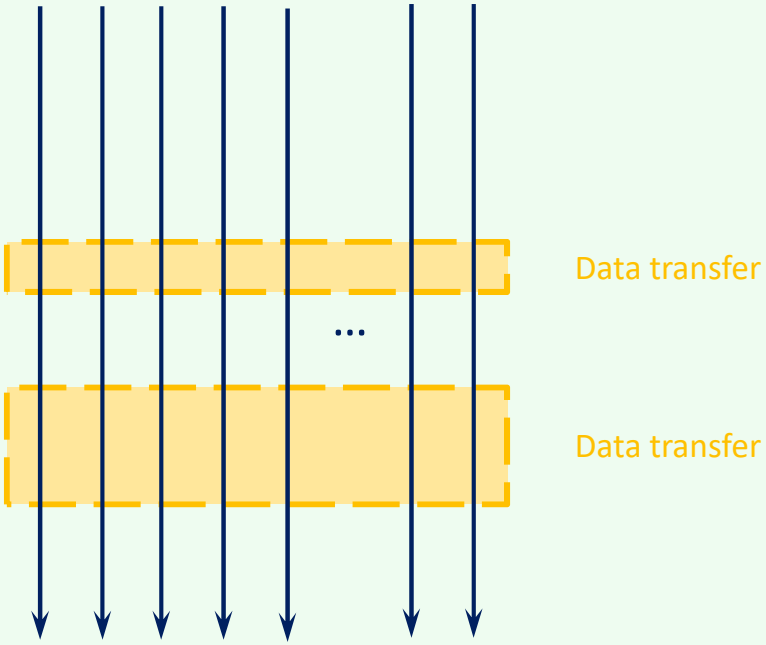
# *Supercomputing Co-Design*



# *No Overlapping Between Computation and Communication Areas*

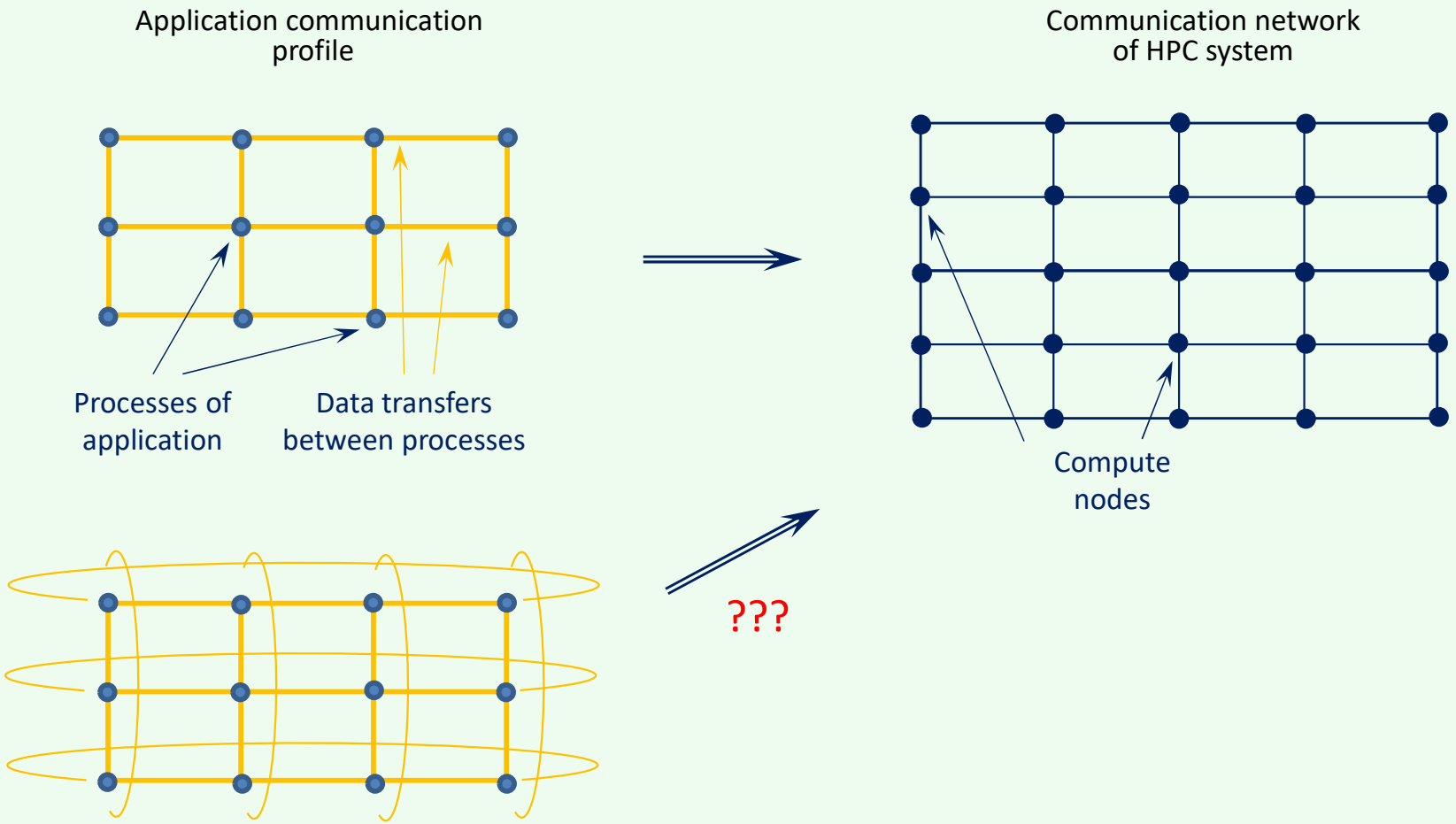



Original version

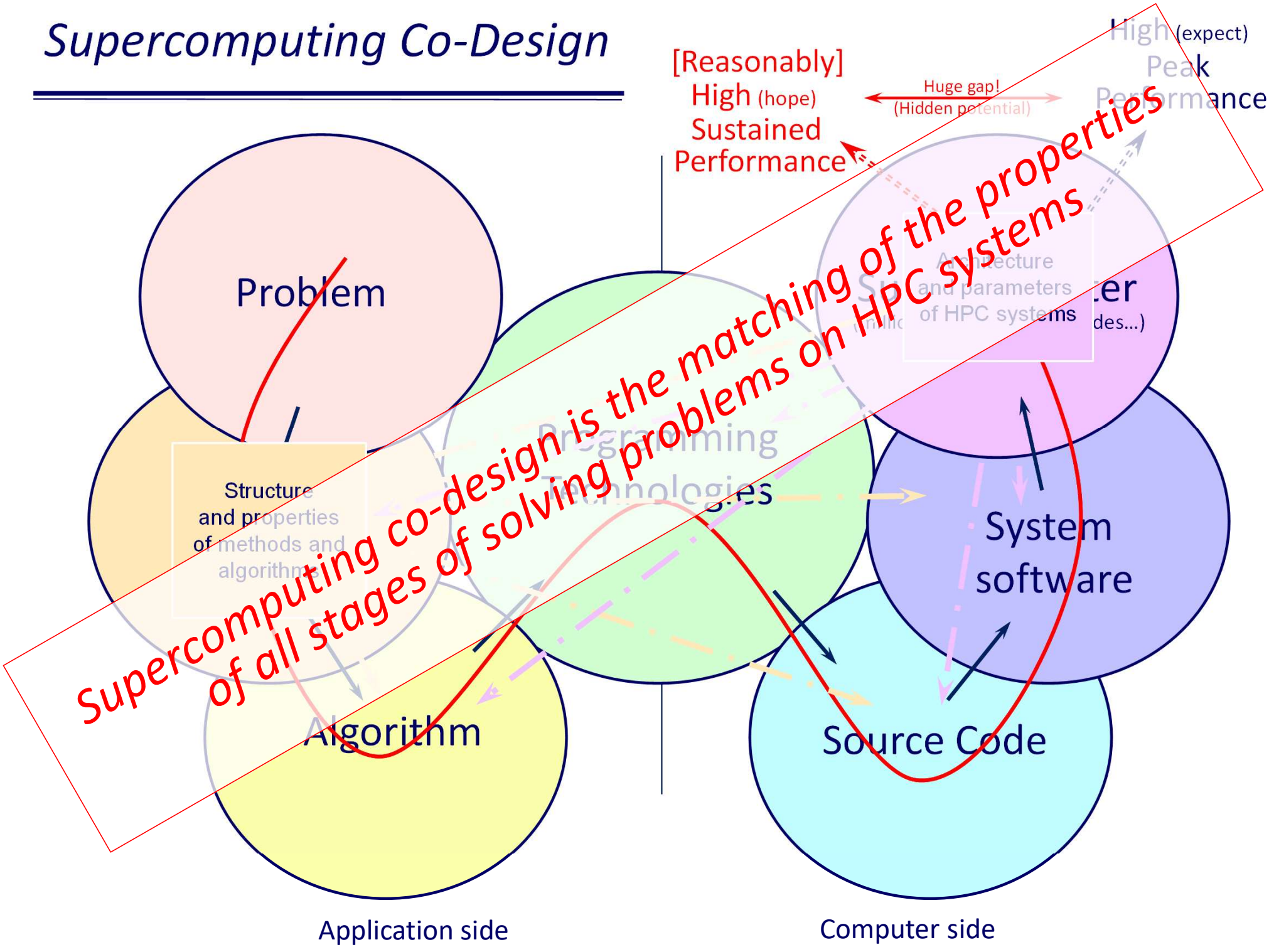



Most likely a more efficient version

# *Mismatch of Communication Network Topology and Application Communication Profile*



# Supercomputing Co-Design



# *Supercomputer Centers Today*

- Huge potential,
- Very high cost,
- Technological complexity,
- Difficult support and maintenance,
- Diversity of architectures,
- High performance via complexity of architectures.

# *Supercomputer Centers Today*
## *(MSU Supercomputer "Lomonosov" as an example)*

What does "efficiency of supercomputers" mean? I need to know (control) everything…

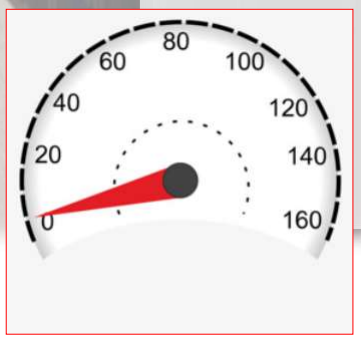
| | |
|---|---|
| Peak Performance | 1.7 Pflops |
| Linpack Performance | 901.9 Tflops |
| Efficiency | 53 % |
| Compute nodes (Intel, x86) | 5 104 |
| Compute nodes (GPU) | 1 065 |
| Intel Xeon (X5570/X5670) | 12 346 |
| GPU (NVIDIA X2070) | 2 130 |
| X86 Cores, total | 52 168 |
| GPU Cores, total | 954 240 |
| RAM | 92 TBytes |
| Interconnect | QDR 4x Infiniband / 10 GE |
| Data Storage System | 1.75 Pbytes, Lustre, NFS, … |
| Operating System | Clustrx T-Platforms Edition |
| Footprint (supercomputer) | 252 $m^2$ |
| Power Consumption | 2.7 MW |
| Launch Year (Last Upgrade) | 2009 (2012) |

Software stack components

Deep memory hierarchy

Users

Projects




← special actions are desperately needed to prevent almost-zero speed/efficiency !

# *Key Components of Intelligent Control*
## *(MSU Supercomputing Center)*

*We must control everything what is necessary to control efficiency permanently and exactly.*

*We need guarantee of coincidence between our expectations and reality.*


*We must describe everything that needs to be controlled.*

# TASC – detected issues, examples

## Basic information

| | |
|---|---|
| Cluster | lomonosov-2 |
| Job ID | |
| Login | |
| State | COMPLETED |
| Partition | compute |
| Num cores | 336 |
| Num nodes | 24 |
| Submit time | 10/11/18 18:02:15 |
| Start time | 10/11/18 19:35:17 |
| End time | 10/12/18 05:20:15 |
| Duration (hours) | 9.7 |

Exec path: /home/

## Thresholds

| Metric | Value | Ranking |
|---|---|---|
| Average CPU load (%) | 98.57 | good |
| Average Loadavg | 1819.40 | low |
| Average IPC | 1.15 | good |
| Average GPU load (%) | 0.00 | low |
| Average IB receive data MPI (MB/s) | 0.00 | low |
| Average IB send data MPI (MB/s) | 0.00 | low |
| Average IB receive data FS (MB/s) | 0.00 | low |

## Features

- Job has few MPI communications.
- Too low CPU load for current value of load average.
- Job shows intensive memory usage.
- Job has low memory access locality.

*Too much!*

**Abnormal job behavior**

## Detected performance issues

| Type | Description | Supposition | Recommendation |
|---|---|---|---|
| | Job is running in the partition for GPU-based programs but almost do not use graphical processors. | The partition for this job is selected incorrectly. | It is recommended to change the partition for this job. |
| | The job shows abnormally low efficiency. | The job is hanged or working incorrectly. | It is recommended to check if the job is launched correctly and cancel it if necessary. |

# *TASC – detected issues, examples*

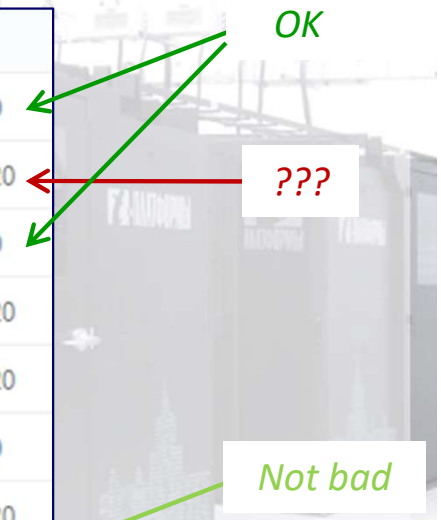
## Abnormal everything

| | Detected issues | Job ID | Start time | End time | End status | # of nodes | Duration, hours | Job size (CPU*h) | CPU load | GPU load | Load average | IPC | Bytes received via MPI, MB/sec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ≡ | | | 2019-04-12 19:19:55 | 2019-04-12 19:36:35 | completed | 1 | 0.3 | 3.3 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| ≡ | | | 2019-04-18 21:36:17 | 2019-04-18 23:47:05 | completed | 1 | 2.2 | 26.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ≡ | | | 2019-04-18 21:40:53 | 2019-04-18 23:46:58 | completed | 1 | 2.1 | 29.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ≡ | | | 2019-04-18 23:53:53 | 2019-04-19 02:07:40 | completed | 1 | 2.2 | 31.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ≡ | | | 2019-04-19 09:35:32 | 2019-04-19 15:38:09 | cancelled | 1 | 6.0 | 84.6 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| ≡ | | | 2019-04-19 15:39:19 | 2019-04-19 17:16:02 | cancelled | 1 | 1.6 | 22.6 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| ≡ | | | 2019-04-19 17:16:07 | 2019-04-19 17:32:12 | completed | 1 | 0.3 | 3.8 | 1.5 | 0.0 | 0.1 | 0.0 | 0.0 |

# *TASC: fine properties of user+package*

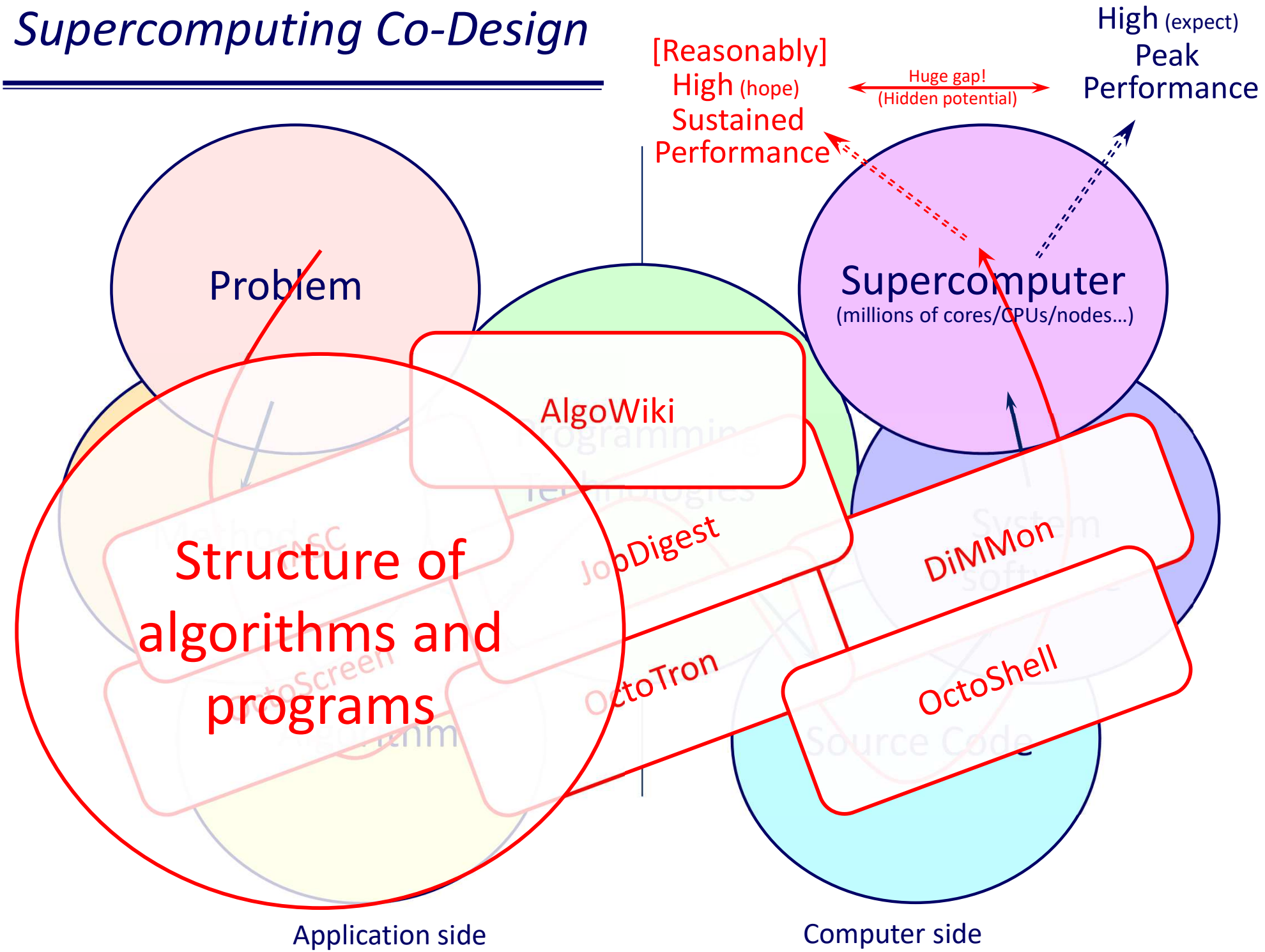
- Gromacs user data:
  - Processor load is high;
  - There are no problems with network and memory usage;
  - But GPU accelerators are not involved at all.

| name | value | ranking |
|---|---|---|
| CPU load | 88.6 | № 3 / 20 |
| GPU load | 0.0 | № 19 / 20 |
| Load average | 24.9 | № 3 / 20 |
| IPC (instructions per second) | 0.8 | № 17 / 20 |
| Intensity of FS usage (bytes sent/received per second) | 9.2k | № 16 / 20 |
| Intensity of memory usage (load/store operations per second) | 764.1m | № 8 / 20 |
| Intensity of MPI usage (bytes received per second) | 339.4m | № 11 / 20 |
| Amount of L1 cache misses per second | 26.5m | № 3 / 20 |
| Amount of LLC cache misses per second | 84.3k | № 16 / 20 |




- Result of analysis: custom build of Gromacs was used, GPU driver did not support the latest version of CUDA, the user was unaware…
- Correction after analysis: average GPU load increased to 43%.

# *Supercomputing Co-Design*

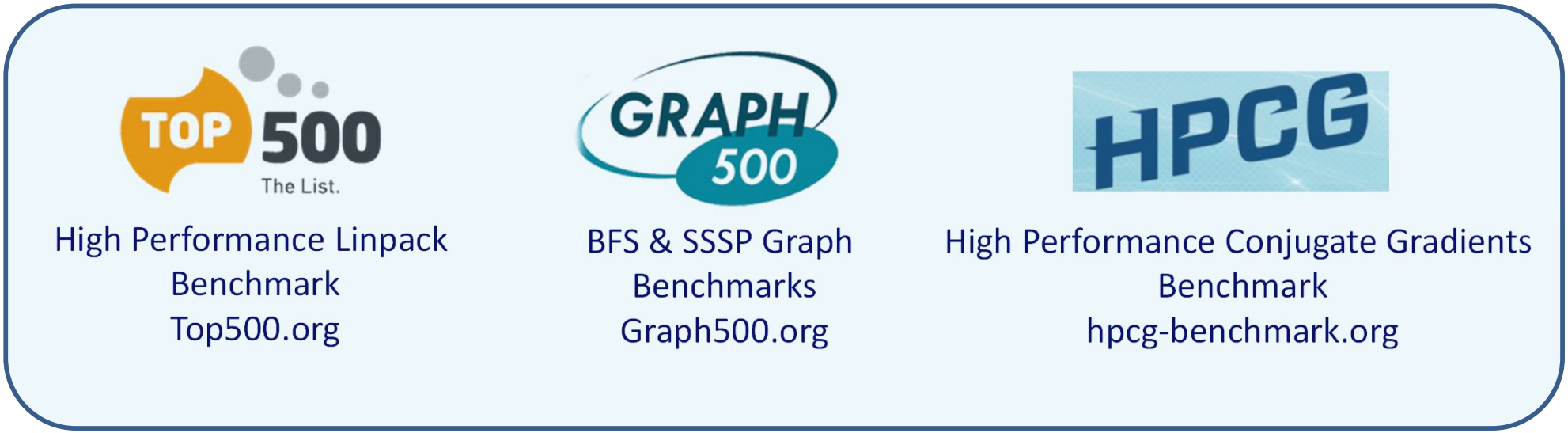


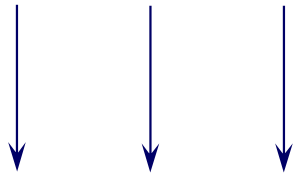
# *Computer architecture landscape today*

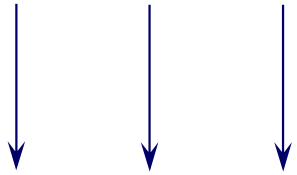
*(How diverse is a computer world today?)*


*We should think about the choice of computer platforms…*

# *Existing methodologies to design computing platforms*
*(Top500, Graph500, HPCG)*



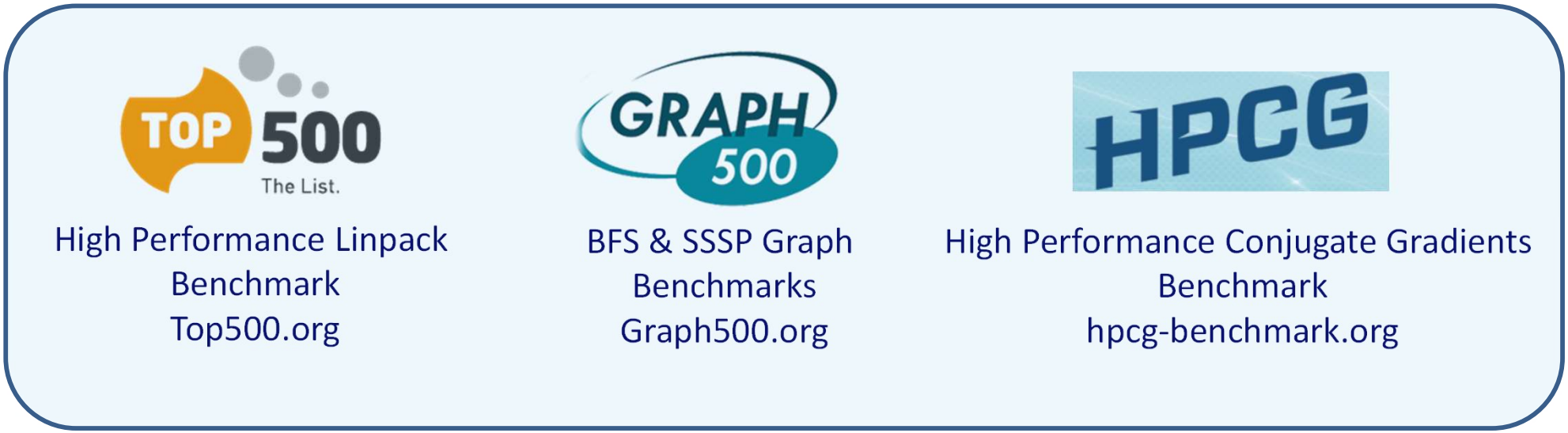



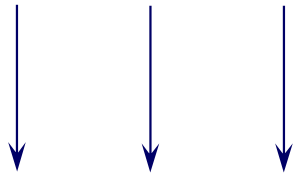


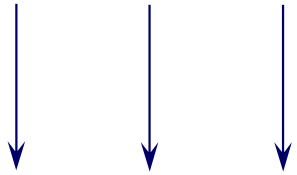
Well-known theoretical potential of algorithms

Well-described and available community experience

# *Existing methodologies to design computing platforms*
*(Top500, Graph500, HPCG)*





Well-known theoretical potential of algorithms

Well-described and available community experience

# *General methodology to design computing platforms ?*

*(oriented to specific algorithms)*

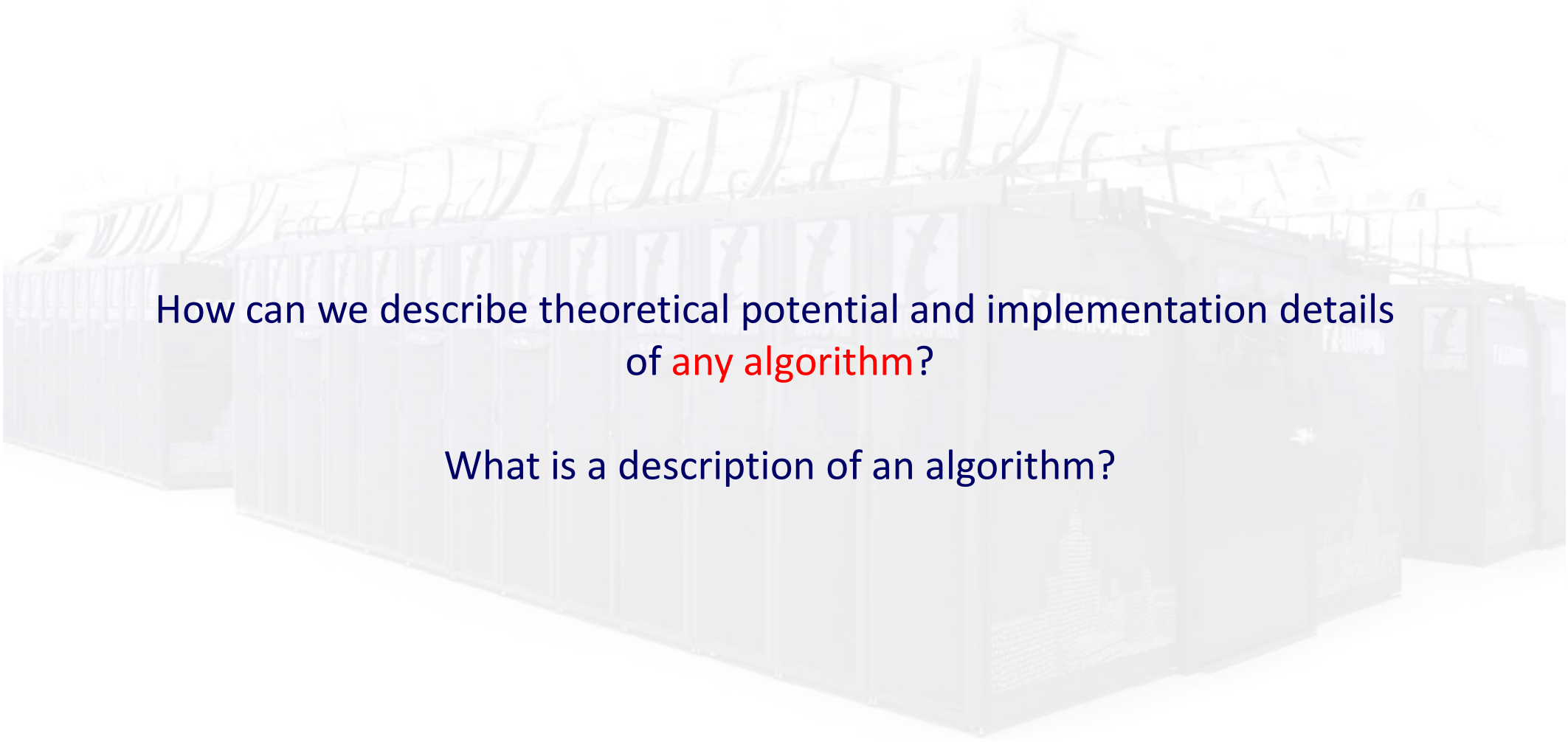
?

Well-known theoretical potential of algorithms

?

Well-described and available community experience

# *“Well-known theoretical potential of algorithms”… How to find it?...*



How can we describe theoretical potential and implementation details of any algorithm?

What is a description of an algorithm?

# *Generations of Parallel Computer Architectures*

*(or How often we had to rewrite our applications completely?)*

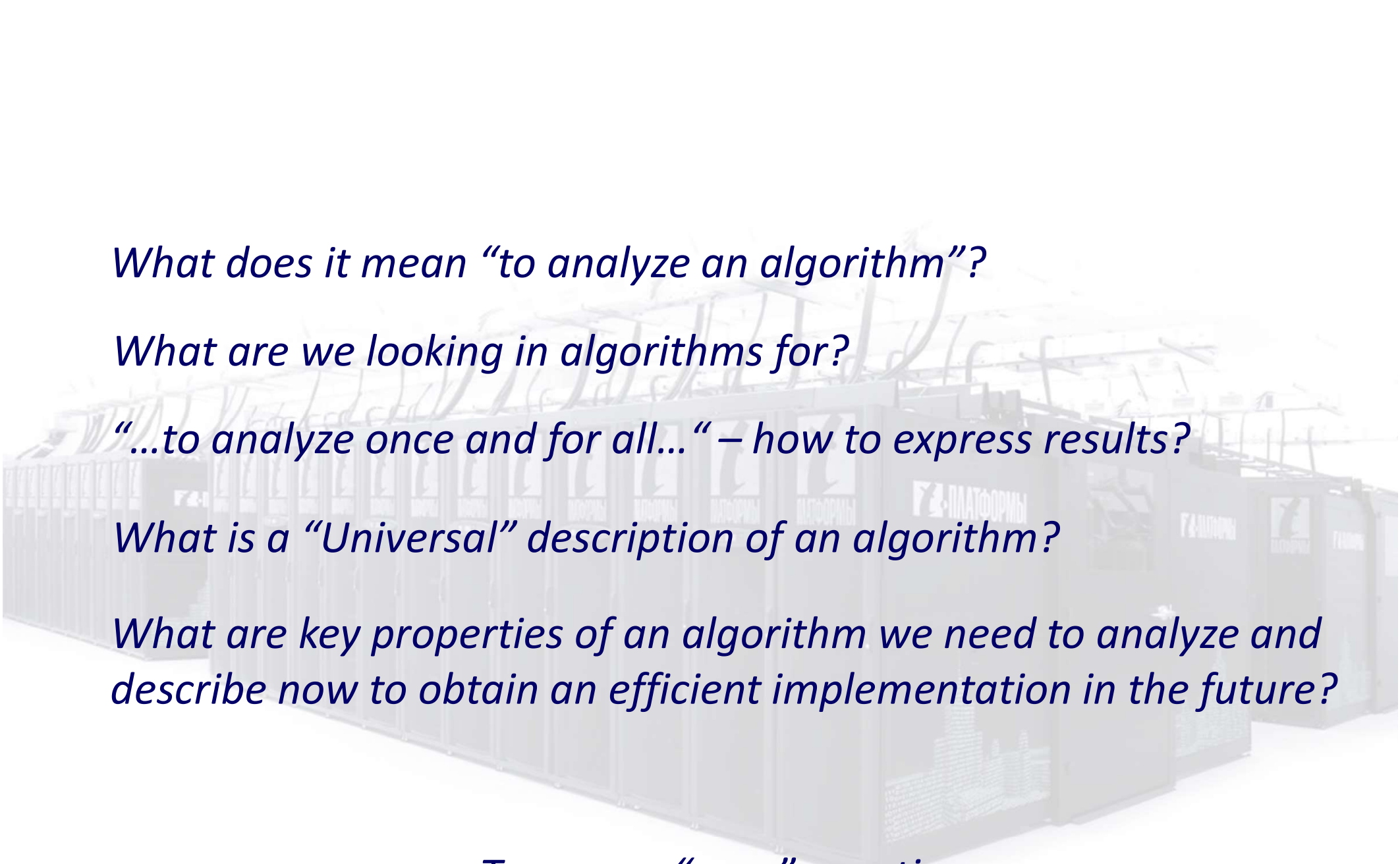
*Parallel programming paradigms (from the 70s up to now):*

*70s - Loop Vectorization (innermost)*
*80s - Loop Parallelization (outer) + Vectorization (innermost)*
*90s - MPI*
*mid 90s - OpenMP*
*mid 2000s - MPI+OpenMP*
*2010s - CUDA, OpenCL, MPI+OpenMP+accelerators (GPU, Xeon Phi)*
*…*

*Can we analyze algorithms once and for all ?*

*Changes in computer architectures do not change algorithms! But…*

*For each generation of a new computing platform we have to:*
*- Analyze algorithms to find a way to match better characteristics of the platform ;*
*- Express the properties of algorithms we found to obtain efficient Implementation for the platform.*



*What does it mean "to analyze an algorithm"?*

*What are we looking in algorithms for?*

*"…to analyze once and for all…" – how to express results?*

*What is a "Universal" description of an algorithm?*

*What are key properties of an algorithm we need to analyze and describe now to obtain an efficient implementation in the future?*

*Too many "easy" questions…*

# *Structure and Properties of Algorithms*

Yes, it can be done: AlgoWiki project

http://AlgoWiki-Project.org/

*Can we analyze algorithms once and for all ?*

*- Analyze algorithms to find a way to match better characteristics of the platform;*

# *The AlgoWiki Project:*
## *description of parallel structure and properties of algorithms*

Open Encyclopedia of Parallel

algowiki-project.org/en/Open_Encyclopedia_of_Parallel_Algorithmic_Features

AlgoWiki

Page | Discussion | Read | View source | View history | Search Algowiki

# Open Encyclopedia of Parallel Algorithmic Features

Main page
Forum
Recent changes

File storage
New files
Upload file

Tools
What links here
Related changes
Special pages
Printable version
Permanent link
Page information

In other languages
Русский

### Welcome! Join us!

**AlgoWiki** is an open encyclopedia of **algorithms' properties** and **features of their implementations** on different hardware and software platforms from mobile to extreme scale, which allows for collaboration with the worldwide computing community on algorithm descriptions.

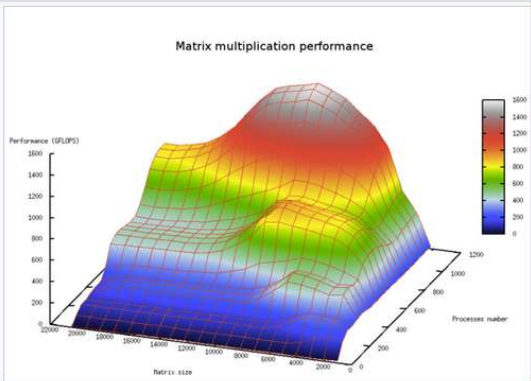
**AlgoWiki** provides an exhaustive description of an algorithm. In addition to classical algorithm properties such as serial complexity, AlgoWiki also presents additional information, which together provides a complete description of the algorithm: its parallel complexity, parallel structure, determinacy, data locality, performance and scalability estimates, communication profiles for specific implementations, and many others.

Read more: About project.

### Project structure

Algorithm classification — the main section of AlgoWiki which contains descriptions of all algorithms. Algorithms are added to the appropriate category of the classification, and classification is expanded with new sections if necessary.

### Today's featured picture




Performance for dense matrix multiplication

### Featured article

**Cholesky decomposition**

1 Properties and structure of the algorithm

**1.1 General description**

**The Cholesky decomposition algorithm** was first proposed by Andre-Louis Cholesky (October 15, 1875 - August 31, 1918) at the end of the First World War shortly before he was killed in battle. He was a French military officer and mathematician. The idea of this algorithm was published in 1924 by his fellow officer and, later, was used by Banachiewicz in 1938 [7]. In the Russian mathematical literature, the Cholesky decomposition is also known as the square-root method [1-3] due to the square root operations used in this

**Properties of the algorithm:**

- Sequential complexity: $O(n^3)$
- Height of the parallel form: $O(n)$
- Width of the parallel form: $O(n^2)$
- Amount of input data: $\frac{n(n+1)}{2}$
- Amount of output data: $\frac{n(n+1)}{2}$

### Work organization

Description of algorithm properties and structure
Guides to writing sections of the algorithm's description
Glossary
Help with editing

### Readiness of articles

**Finished articles**:

- Gaussian elimination, compact scheme for tridiagonal matrices, serial version
- Householder (reflections) method for reducing of a matrix to

*http://AlgoWiki-Project.org*

# *Description of algorithms*

*(What properties of algorithms should be included in the description?)*

For positive definite Hermitian matrices (*symmetric matrices in the real case*), we use the decomposition $A = LL^*$, where $L$ is the lower triangular mat ion $A = U^*U$, where $U$ is the upper triangular matrix. These forms of the Cholesky decomposition are equivalent in the sense of the amount of arithmetic operations and are different in the sense of data represntation. The essence of this decomposition consists in the implementation of formulas obtained uniquely for the elements of the matrix $L$ from features.

General Description

Input data: a symmetric positive definite matrix $A$ whose elements are denoted by $a_{ij}$ ).

Output data: the lower triangular matrix $L$ whose elements are denoted by $l_{ij}$ ).

The Cholesky algorithm can be represented in the fo

$$l_{11} = \sqrt{a_{11}},$$

$$l_{j1} = \frac{a_{j1}}{l_{11}}, \quad j \in [2, n],$$

$$l_{ii} = \sqrt{a_{ii} - \sum_{p=1}^{i-1} l_{ip}^2}, \quad i \in [2, n],$$

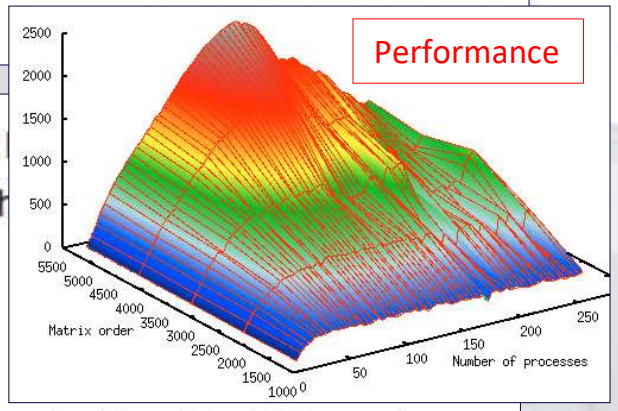
$$l_{ji} = \left( a_{ji} - \sum_{p=1}^{i-1} l_{ip} l_{jp} \right) / l_{ii}, \quad i \in [2, n-1], j \in [i+1, n].$$

Mathematical Description

The following number of operations should matrix of order $n$ using a serial version of th

- $n$ square roots,
- $\frac{n(n-1)}{2}$ divisions,
- $\frac{n^3-n}{6}$ multiplications and $\frac{n^3-n}{6}$ additions (subtractions): the main amount of computational work

Serial Complexity

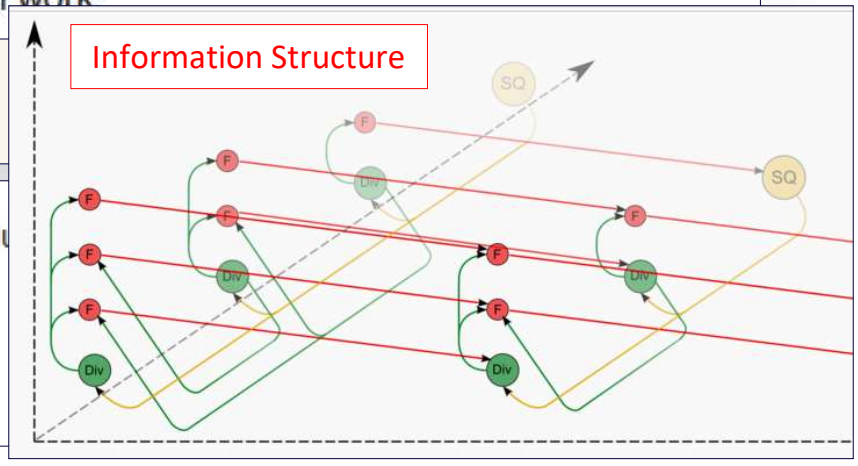



Performance

A computational kernel of its serial version can be composed of $\frac{n(n-1)}{2}$ dot produ

$$\sum_{p=1}^{i-1} l_{ip} l_{jp}.$$

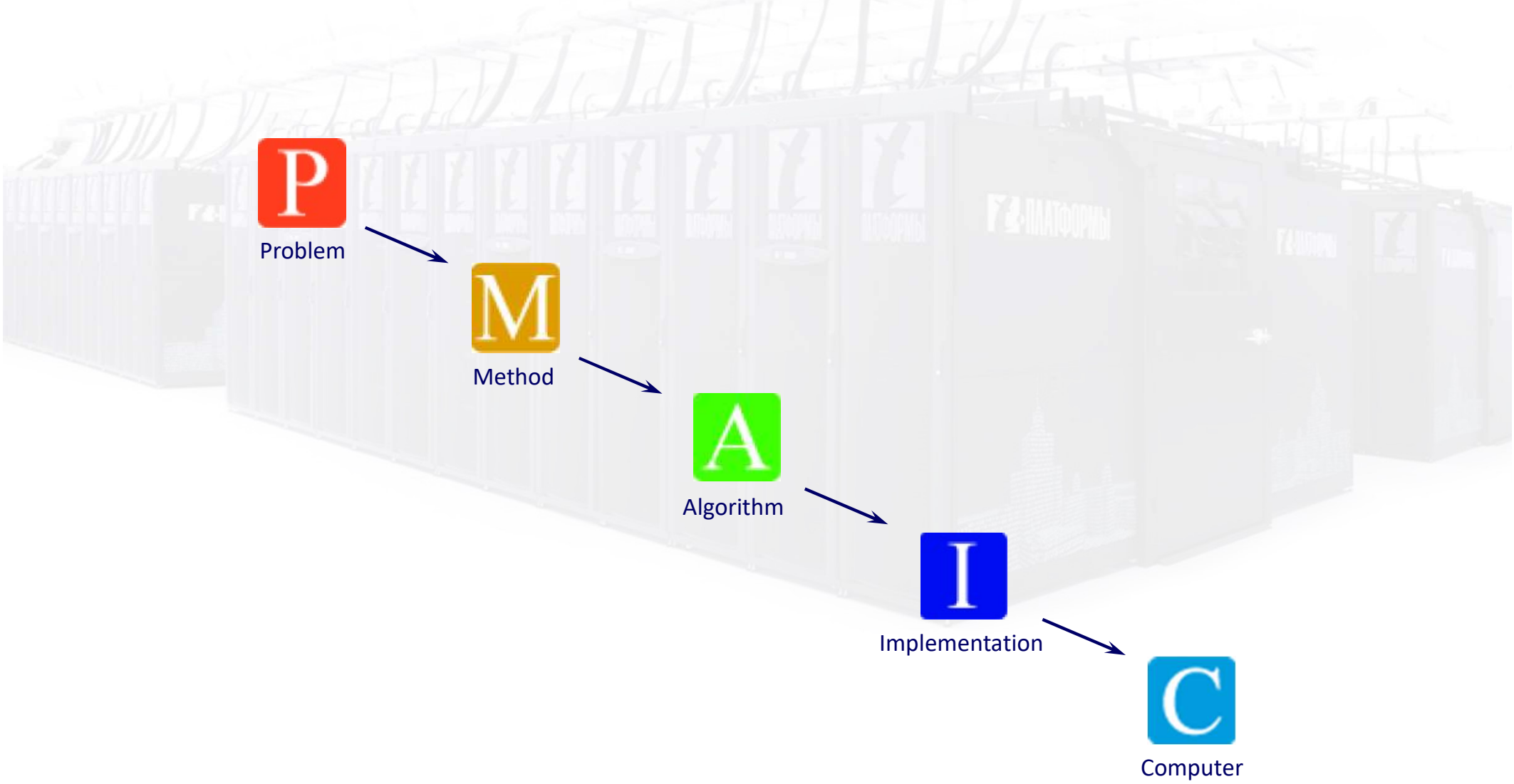
Computational Kernel




Information Structure

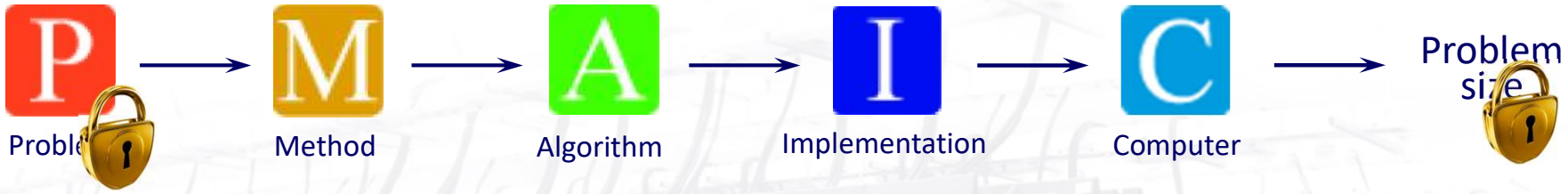
# *AlgoWiki: Problem – Method – Algorithm – Implementation – Computer Platform*

*(What do we have for each algorithm in AlgoWiki?*
*An exhaustive description of the chain)*



# *"PerfData" list on "Single Source Shortest Path"*

*(How to solve the problem of the given size?)*




| Method | Implementation | Computing Platform | MTEPS | GraphType | GraphSize |
|---|---|---|---|---|---|
| Bellman–Ford | RCC for GPU | Lomonosov | 1309,0 | SSCA-2 | 2^20 |
| Delta-Stepping | GAP | Lomonosov-2 | 512,0 | RMAT | 2^20 |
| Bellman–Ford | Ligra | Lomonosov-2 | 511,0 | RMAT | 2^20 |
| Bellman–Ford | RCC for GPU | Lomonosov | 452,9 | SSCA-2 | 2^20 |
| Bellman–Ford | RCC for CPU | Lomonosov-2 | 418,0 | RMAT | 2^20 |
| Bellman–Ford | Graph500 MPI | Lomonosov | 350,0 | RMAT | 2^20 |
| Bellman–Ford | RCC for CPU | Lomonosov | 204,1 | RMAT | 2^20 |
| Bellman–Ford | RCC for CPU | Lomonosov | 183,5 | SSCA-2 | 2^20 |
| Dijkstra's | PBGL MPI | Cluster / "Angara" interconnect | 150,0 | SSCA-2 | 2^20 |
| Bellman–Ford | Graph500 MPI | Lomonosov | 120,0 | RMAT | 2^20 |
| Bellman–Ford | Graph500 MPI | Lomonosov | 18,0 | SSCA-2 | 2^20 |
| Dijkstra's | PBGL MPI | IBM BlueGene/P | 8,9 | SSCA-2 | 2^20 |
| Delta-Stepping | PBGL MPI | IBM BlueGene/P | 3,8 | SSCA-2 | 2^20 |
| Delta-Stepping | PBGL MPI | IBM BlueGene/P | 1,3 | RMAT | 2^20 |
| Dijkstra's | PBGL MPI | IBM BlueGene/P | 0,6 | RMAT | 2^20 |

# *AlgoWiki: an easy step back to analyze "PerfData"*
## *(theoretical potential of algorithms is described in AlgoWiki)*

SSSP problem:

| Method | Implementation | Computing Platform | MTEPS | GraphType | GraphSize |
|---|---|---|---|---|---|
| Bellman–Ford | RCC for CPU | Lomonosov-2 | 418,0 | RMAT | 2^20 |
| Bellman–Ford | Graph500 MPI | Lomonosov | 350,0 | RMAT | 2^20 |
| Bellman–Ford | RCC for CPU | Lomonosov-2 | 204,1 | RMAT | 2^20 |
| Dijkstra's | PBGL MPI | Cluster / "Angara" interconnect | 150,0 | SSCA-2 | 2^20 |
| Delta-Stepping | PBGL MPI | Lomonosov | 124,1 | SSCA-2 | 2^21 |
| Bellman–Ford | Graph500 MPI | Lomonosov | 120,0 | RMAT | 2^20 |
| Dijkstra's | PBGL MPI | IBM BlueGene/P | 8,9 | SSCA-2 | 2^20 |
| Dijkstra's | PBGL MPI | Lomonosov | 5,3 | SSCA-2 | 2^21 |
| Delta-Stepping | PBGL MPI | IBM BlueGene/P | 3,8 | SSCA-2 | 2^20 |


AlgoWiki

Page | Discussion

### Dijkstra's algorithm

Main page
Forum
Recent changes
File storage
New files
Upload file
Tools
What links here
Related changes
Special pages
Printable version

Primary authors of this description: A.N.Daryin, Vad.V.Voevodin (Section 2.2).

Contents [hide]

1 Properties and structure of the algorithm
1.1 General description of the algorithm
1.2 Mathematical description of the algorithm
1.3 Computational kernel of the algorithm
1.4 Macro structure of the algorithm
1.5 Implementation scheme of the serial algorithm
1.6 Serial complexity of the algorithm

**1.6 Serial complexity of the algorithm**

The serial complexity of the algorithm is $O(C_1 m + C_2 n)$, where

- $C_1$ is the number of operations for decreasing the distance to a node;
- $C_2$ is the number of operations for calculating minima.

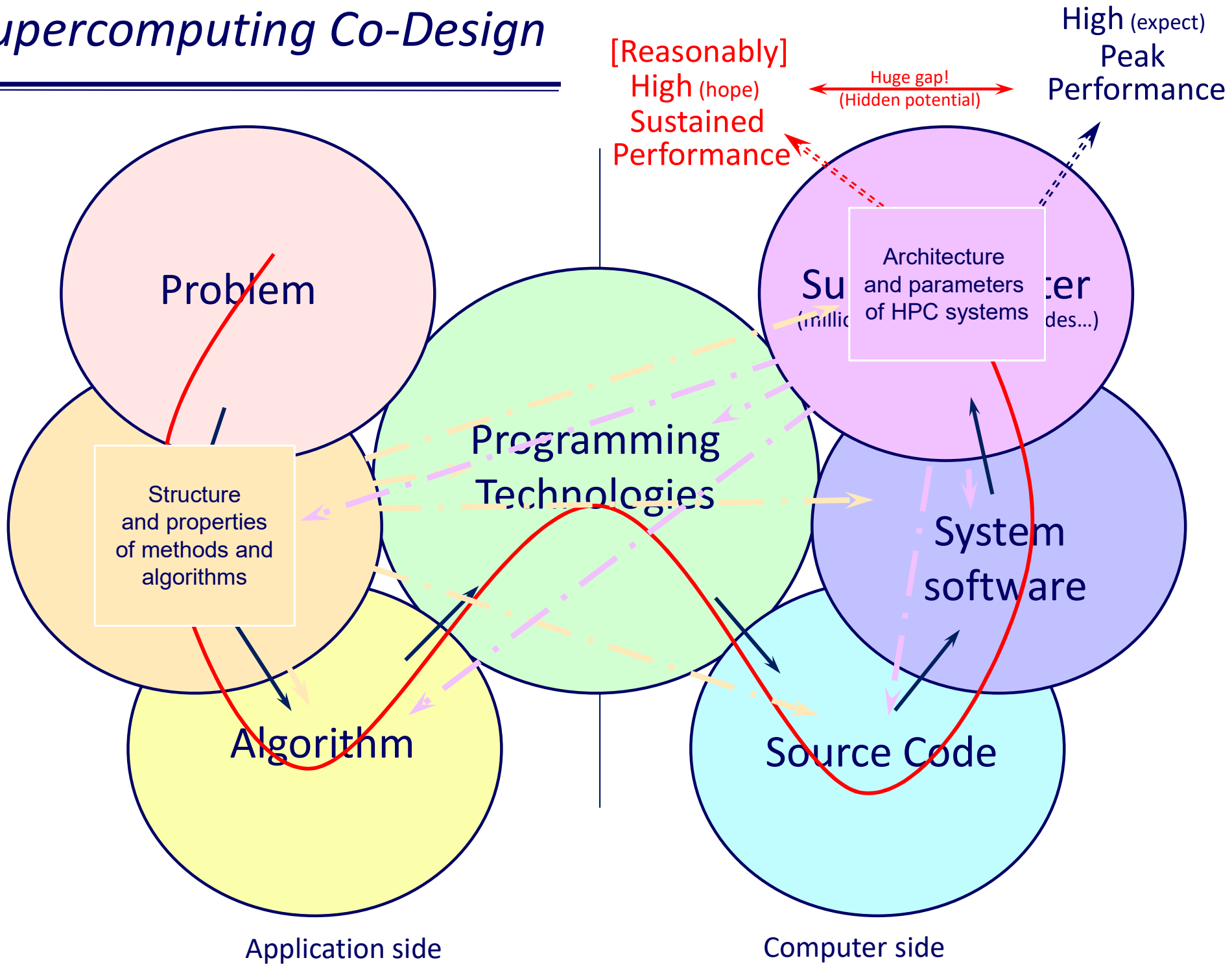
The original Dijkstra's algorithm used lists as an internal data structure. For such lists, $C_1 = O(1)$, $C_2 = O(n)$, and the total complexity is $O(n^2)$.

**1.8 Parallelization resource of the algorithm**

Dijkstra's algorithm admits an efficient parallelization [3] Its average execution time is $O(n^{1/3} \ln n)$, and the computational complexity is $O(n \ln n + m)$.

The algorithm of Δ-stepping can be regarded as a parallel version of Dijkstra's algorithm.

# *Supercomputing Co-Design*



*There are numerous ways to apply supercomputing co-design technologies on practice…*

*for instance, designing an HPC system for solving a particular class of problems:*

*There are numerous ways to apply supercomputing co-design technologies on practice…*

*for instance, designing an HPC system for solving a particular class of problems:*

*There are numerous ways to apply supercomputing co-design technologies on practice…*

*for instance, designing an HPC system for solving a particular class of problems:*

Architecture and parameters of HPC system that match properties of algorithms for the given class of problems

# *Algorithm Description Structure*
## *(upper level)*

1. General description
2. Mathematical description
3. Computational kernel
   (operations, data format, data structures)
4. Macrostructure of algorithm
   (macrooperations, typical algorithmic structures)
5. General scheme of serial algorithm
6. Complexity of serial algorithm
   (arithmetic operations, read/write operations)
7. Information graph
8. Resource of parallelism
   (complexity of parallel algorithm, available parallelism)
9. Input and output data
10. Properties of algorithm
    (computational intensity, stability, determinism, imbalance…)
11. Data locality, locality of computations
12. Possible approaches to parallel implementations
    (features, properties, programming technologies…)
13. Possible obstacles for scalability
    (limited resource of parallelism, load imbalance, synchronizations…)
14. Dynamic properties and parameters: features of parallel implementations
15. Existing implementations

# *Description structure of an HPC system*
## *(upper level)*

1. Configuration of HPC system as a whole
2. Parameters of partitions
3. Parameters of computing nodes
4. Parameters of data storage system

---

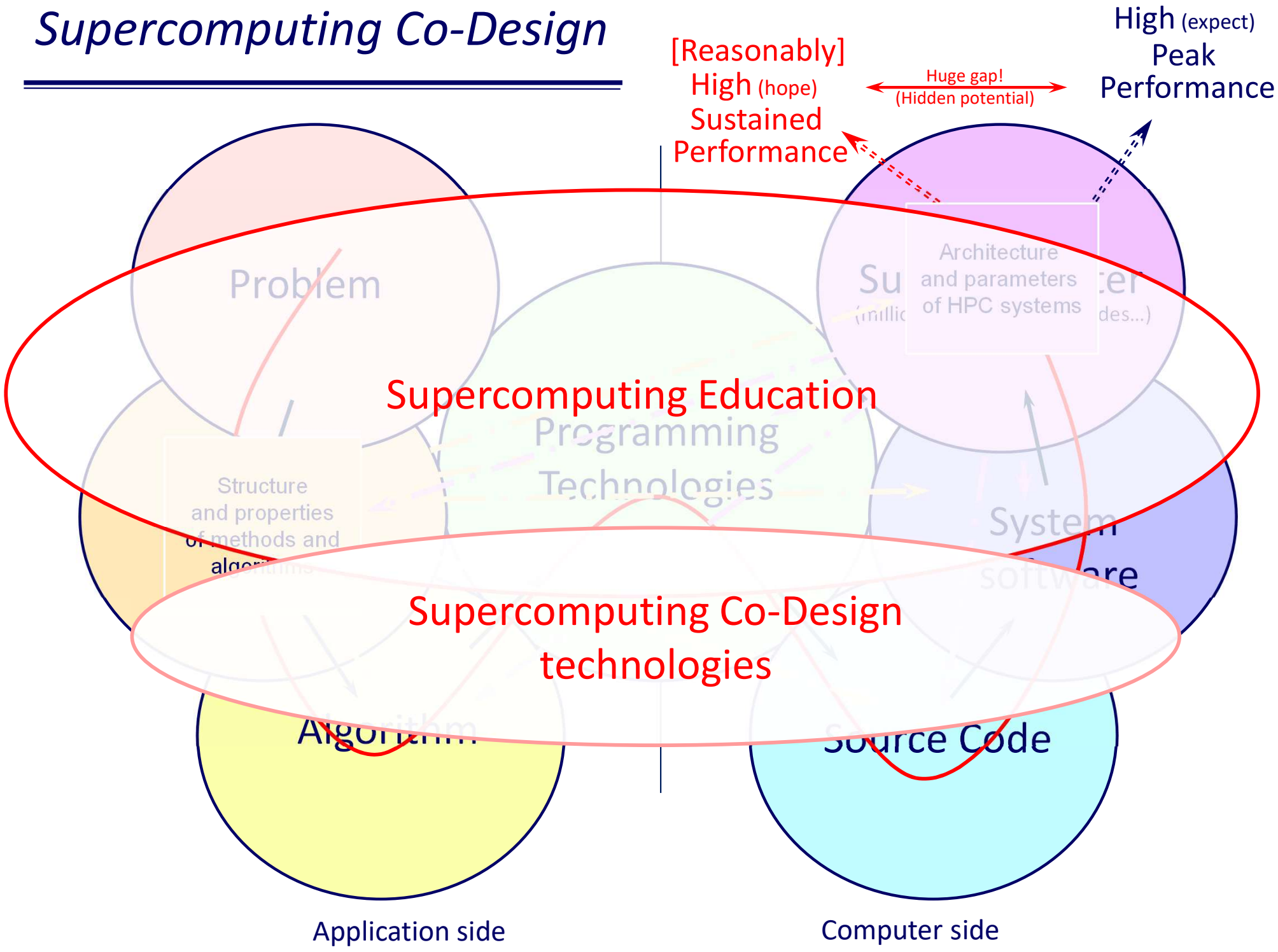
1. Configuration of HPC system as a whole
   - 1.1. Number of partitions in HPC system
   - 1.2. Communication technology between partitions
   - 1.3. Total number of computing nodes
   - 1.4. Total number of cores
   - 1.5. Total amount of memory
2. Parameters of partitions
   - 2.1. Number of computing nodes in partition
   - 2.2. Number of cores in partition
   - 2.3. Amount of memory in partition

     **Communication network**
   - 2.4. Topology
   - 2.5. Latency
   - 2.6. Bandwidth
   - 2.7. Average distance between nodes
   - 2.8. Number of immediate neighbors of each node
3. Parameters of computing nodes
   - 3.1. Number of host processors in a node
   - 3.2. Number of cores in a host processor
   - 3.3. Total number of cores within each node
   - 3.4. Hyperthreading in the host processor
   - 3.5. Vector data processing features in a core

     **Memory structure and hierarchy**
   - 3.6. Amount of memory in a node
   - 3.7. Amount of memory per core
   - 3.8. Difference in data access time to different areas of memory
   - 3.9. Number of cache memory levels
   - 3.10. Size of cache memory for different levels L{1,2,3…}
   - 3.11. Latency of cache memory for different levels L{1,2,3…}
   - 3.12. Latency of the main memory
   - 3.13. Processor-memory channel bandwidth
   - 3.14. Number of channels to memory
   - 3.15. Fast memory (HBM)

     **Accelerators**
   - 3.16. Number of accelerators per node
   - 3.17. Type of accelerators (GPU, FPGA,…)
   - 3.18. Processor-GPU channel bandwidth
   - 3.19. Amount of memory per accelerator
   - 3.20. Communication technology between accelerators
4. Parameters of data storage system

# *Supercomputing Co-Design*




МГУ САРОВ
ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР ФИЗИКИ И МАТЕМАТИКИ
Филиал МГУ имени М. В. Ломоносова

Sarov.msu.ru

СУПЕРКОМПЬЮТЕРНЫЕ ДНИ В РОССИИ

О КОНГРЕССЕ О КОНФЕРЕНЦИИ ПРОГРАММА УЧАСТИЕ НОВОСТИ

# Суперкомпьютерные дни в России 2023

международная научная конференция

25 - 26 сентября

два дня — множество параллельных событий

Семинар в рамках конференции:
«Высшее образование
для цифрового будущего:
конвергенция HPC, Big Data, ML и IoT»

Премия
имени В.П.Гергеля
за лучшие молодежные доклады

Научные школы

Суперкомпьютерная Академия

23.09 – 30.09.2023

*10th International Conference*
*"Distributed Computing and GRID Technologies in Science and Education"*

# *Thank you for your attention!*

*voevodin@parallel.ru*

*July, 3rd, 2023, JINR, Dubna*