NATIONAL RESEARCH UNIVERSITY

Суперкомпьютерные дни в России 2023

National Research University Higher School of Economics

Supercomputer Atomistic Modelling and Multi-scale Analysis

GPU-ACCELERATED MATRIX EXPONENT FOR SOLVING 1D TIME-DEPENDENT SCHRODINGER EQUATION

Choi Yea Rem, Stegailov V.V.

Moscow, 2023

Contents

TIME DEPENDENT SCHRODINGER EQUATION

Time-Dependent Schrodinger Equation (TDSE)

Fig. 2 The finite-difference real-space approach. For open boundary conditions (shown in the figure), wave functions are sampled on a uniform grid within a spherical domain and vanish outside its boundary. Black dots denote grid points, and the blue balls denote atoms (a dimer example is shown)

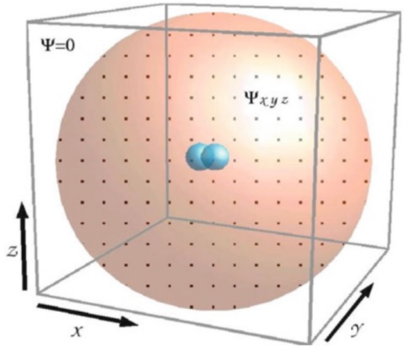
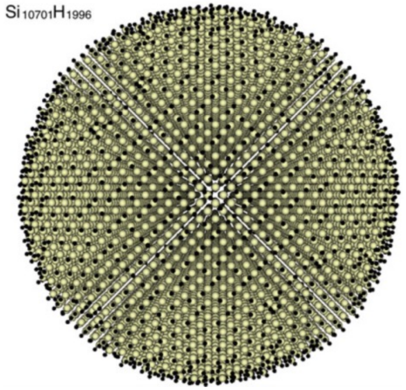


Fig. 7 $Si_{10701}H_{1996}$ nanocrystal the surface is capped with hydrogen atoms shown as a black dots



$$i\hbar\frac{\partial\psi}{\partial t} = H(t)\psi$$

Supercomputer «cHARISMa» HSE University

In one node with 8 A100
Peak 156 TFLOPs, FP32 (Tensor Core)

Chelikowsky, James R. "Extending the Scale with Real-Space Methods for the Electronic Structure Problem." *Handbook of Materials Modeling: Methods: Theory and Modeling* (2020): 499-522.

Algorithm of 1D
Time-Dependent Schrodinger Equation (TDSE)

$$i\hbar\frac{\partial}{\partial t}\psi_i = H\psi_i$$

$$i\hbar\frac{\partial \psi}{\partial t} = H(t)\psi$$

$$\psi(r, t+\Delta t) = \exp\left[i(T+V)\Delta t\right]\psi(r,t)$$

V=V(t) !

Taylor series of matrix exponent

A^k

$$\exp A = I + \frac{1}{1!}A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$$

A^{k-1} $\quad \frac{1}{k}$

GEMM

C = α (AB) + β C

$\alpha = \frac{1}{k}$

$\beta = 0$

$$A^k = \frac{1}{k}A^{k-1}A$$

Multi-GPU GEMM

2020 Global Smart Industry Conference (GloSIC)

Matrix-Matrix Multiplication Using Multiple GPUs Connected by Nvlink

Yea Rem Choi
National Research University Higher School of Economics
Moscow, Russia
yerem5884@gmail.com

Vsevolod Nikolskiy
National Research University Higher School of Economics
Moscow, Russia
thevsevak@gmail.com

Vladimir Stegailov
Joint Institute for High Temperatures of Russian Academy of Sciences
Dolgoprudny, Russia
stegailov@gmail.com

***Abstract*—In this work we present an original GPU-only parallel matrix-matrix multiplication algorithm ($C = \alpha A * B + \beta C$) for servers with multiple GPUs connected by NVLink. The algorithm is implemented using CUDA. The data transfer patterns, the communication and computation overlap, and the overall performance of the algorithm are considered. By regulating the commands call order and the sizes of tiles, we tune the uninterrupted asynchronous data transmission and kernel execution. Two cases are considered: when all the data are stored in one GPU and when the matrices are distributed among several GPUs. The execution efficiency of this new algorithm is**

connection between CPU and GPU is a unique feature available in IBM Power CPUs only. A much more widespread type of multi-GPU servers is represented by Nvidia DGX servers that are based on 8-16 GPUs interconnected via NVLink with PCIe connections between GPUs and CPUs.

Highest levels of computational performance of GPUs and ultrahigh bandwidth and low latency of NVLink make such multi-GPU systems a very attractive option for the development of novel high performance computing algorithms. The algorithms for mathematical modelling that perform all

Faster than cuBLASXT!

Choi Y. R., Nikolskiy V., Stegailov V. Matrix-Matrix Multiplication Using Multiple GPUs Connected by Nvlink, in: *2020 Global Smart Industry Conference (GloSIC)*. IEEE, 2020. P. 354-361.

Algorithm of TDSE numerical solution

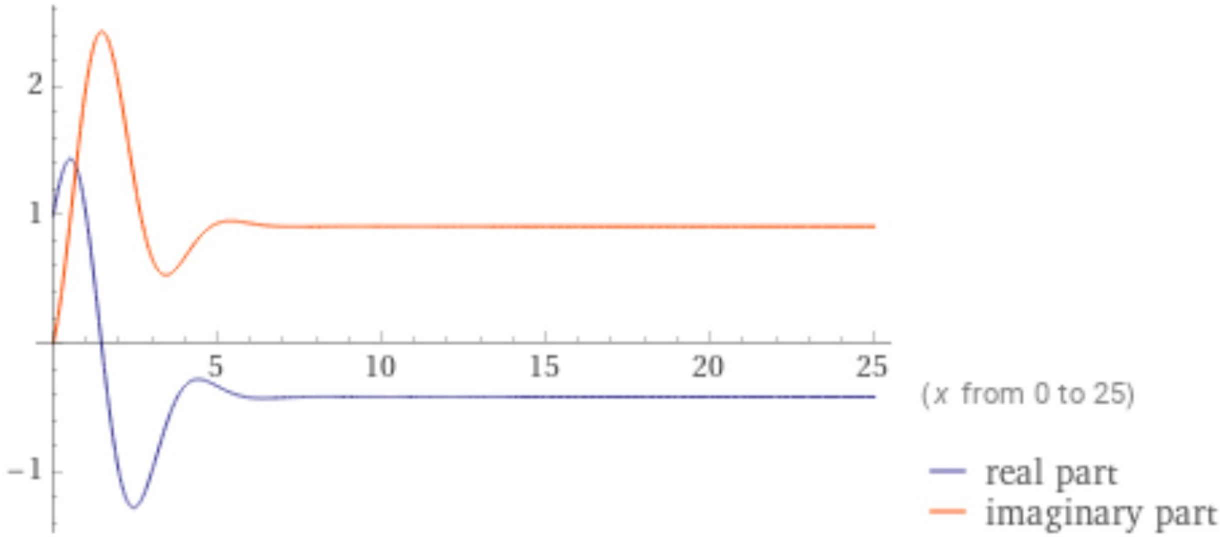
receive (potential);
build Hamiltonian matrix (A);
$expA = A + I$;
for $k = 2$ to rank **do**
    $\alpha = 1/k$;
    Multi-GPU GEMM ($A^k = \alpha A A^{k-1}$);
    $expA = expA + A^k$;
end for
GEMV ($\psi_j = expA\ \psi_{j-1}$);
send (ψ_j); //for output in file

Number of iterations and error rate

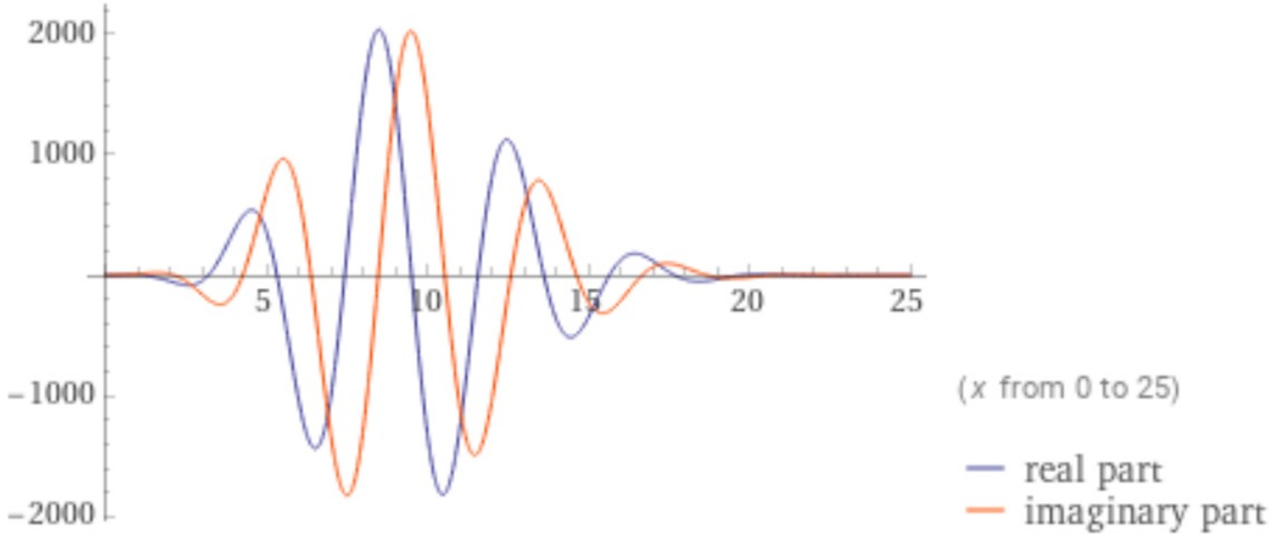
$$\exp iA = \left(I + \frac{i^2}{2!}A^2 + \ldots\right) + \left(\frac{i}{1!}A + \frac{i^3}{3!}A^3 + \ldots\right) = \cos A + i \sin A$$

Number of iterations and error rate

$$\sum_{n=0}^{x} \frac{(2i)^n}{n!}$$



$$\sum_{n=0}^{x} \frac{(10i)^n}{n!}$$



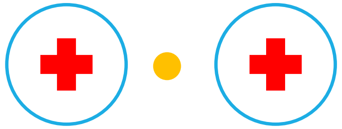
1D TDSE MODEL FOR H_2^+

NUMERICAL EXPERIMENT

Model of the interaction between a hydrogen atom and an electron

Soft-Core Coulomb potential

$$V\ (r) = \frac{-Z}{\sqrt{r^2 + a}}$$



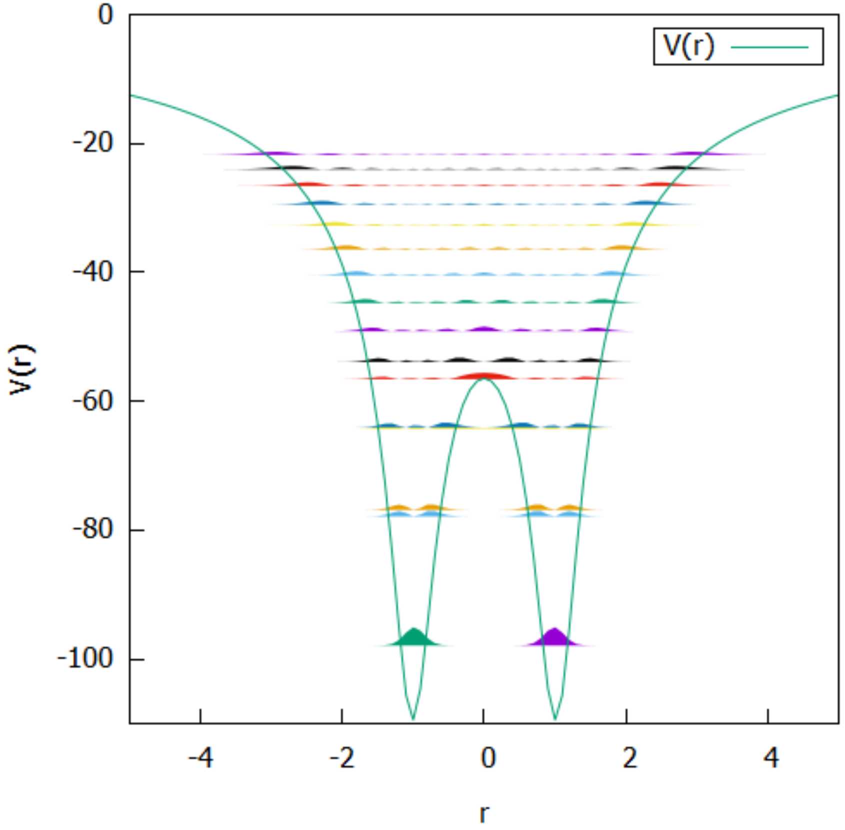
Soft-Core Coulomb potential with trigonometric motion

$$V\ (r, t) = \frac{-Z}{\sqrt{(r\ -\ \alpha \sin(\beta t))^2 + a}}$$

Movement of the center

Solutions of stationary Schrodinger Equation

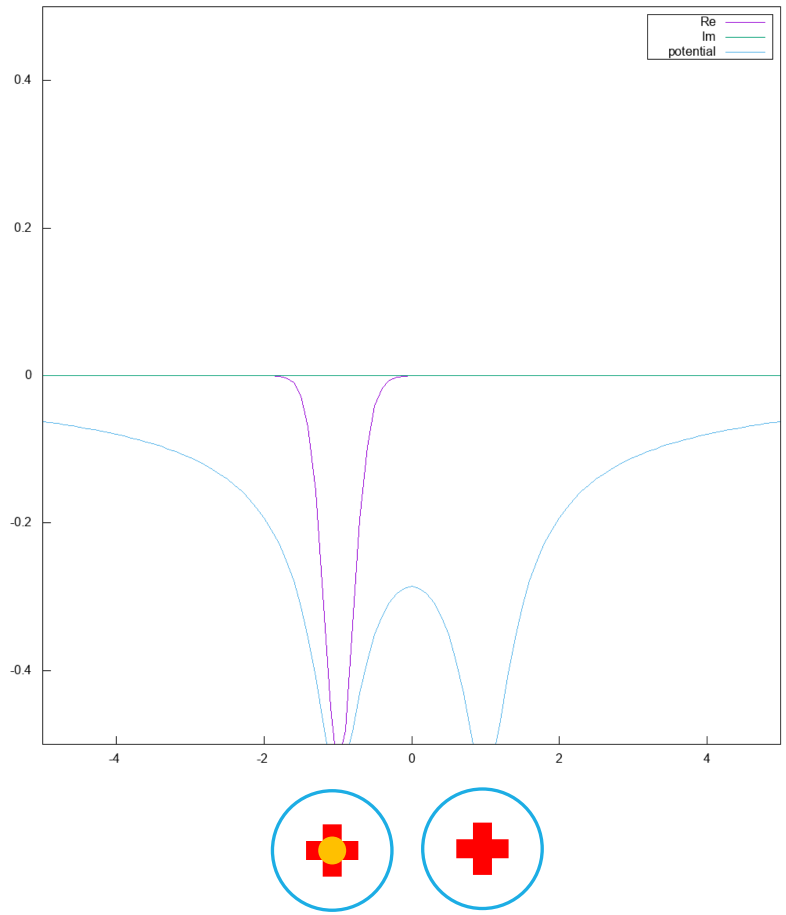
Electron density distributions



$$V\ (r,t) = \frac{-Z}{\sqrt{r^2 + a}}$$

$Z = 30$
$a = 0.1$

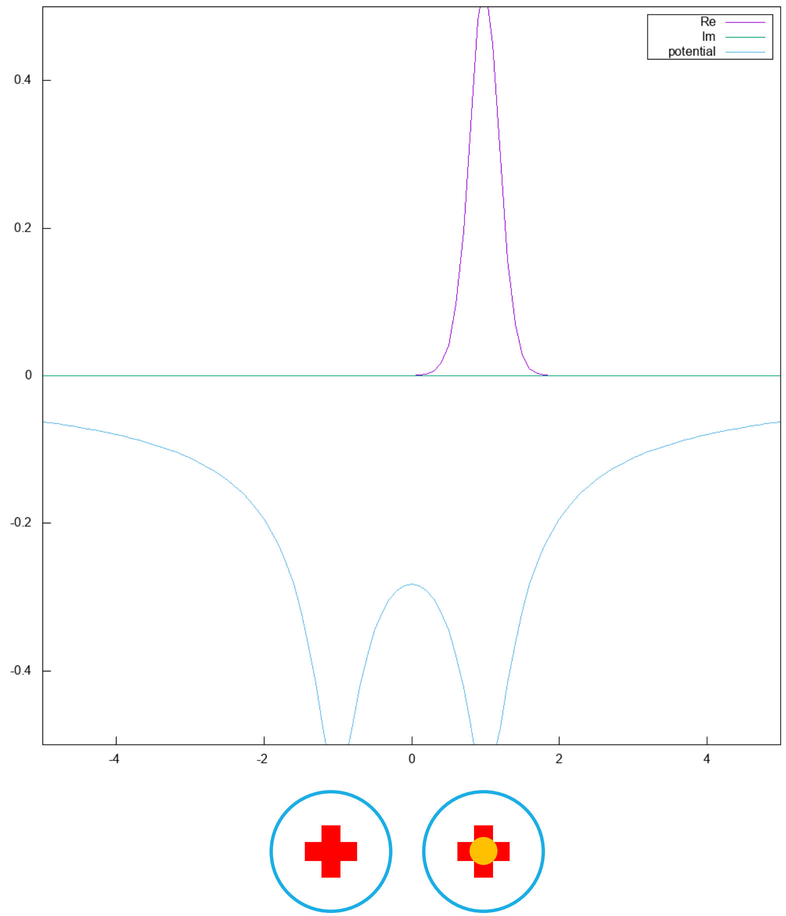
Model with two H^+ and an electron in stationary state



$$V\,(r,t) = \frac{-Z}{\sqrt{r^2 + a}}$$

$Z = 30$
$a = 0.1$
$\Delta r = 0.1$
$\Delta t = 0.002$

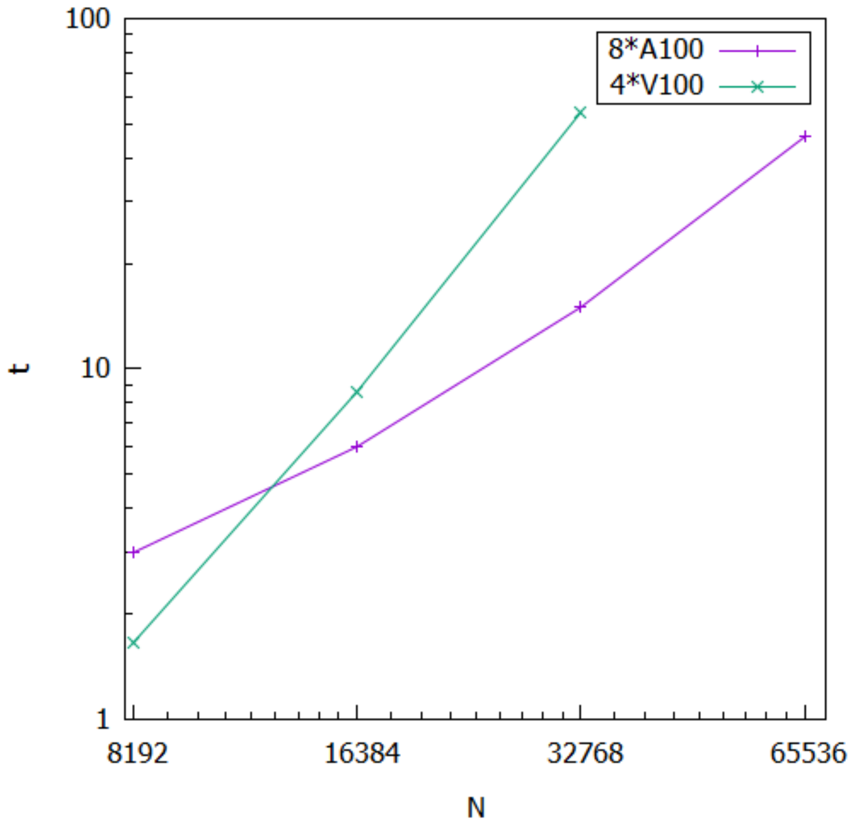
Model with two H^+ and an electron with sinusoidal motion of protons



$$V(r,t) = \frac{-Z}{\sqrt{(r-\alpha\sin(\beta t))^2 + a}}$$

$Z = 30$
$a = 0.1$
$\alpha = 0.5$
$\beta = 3$
$\Delta r = 0.1$
$\Delta t = 0.002$

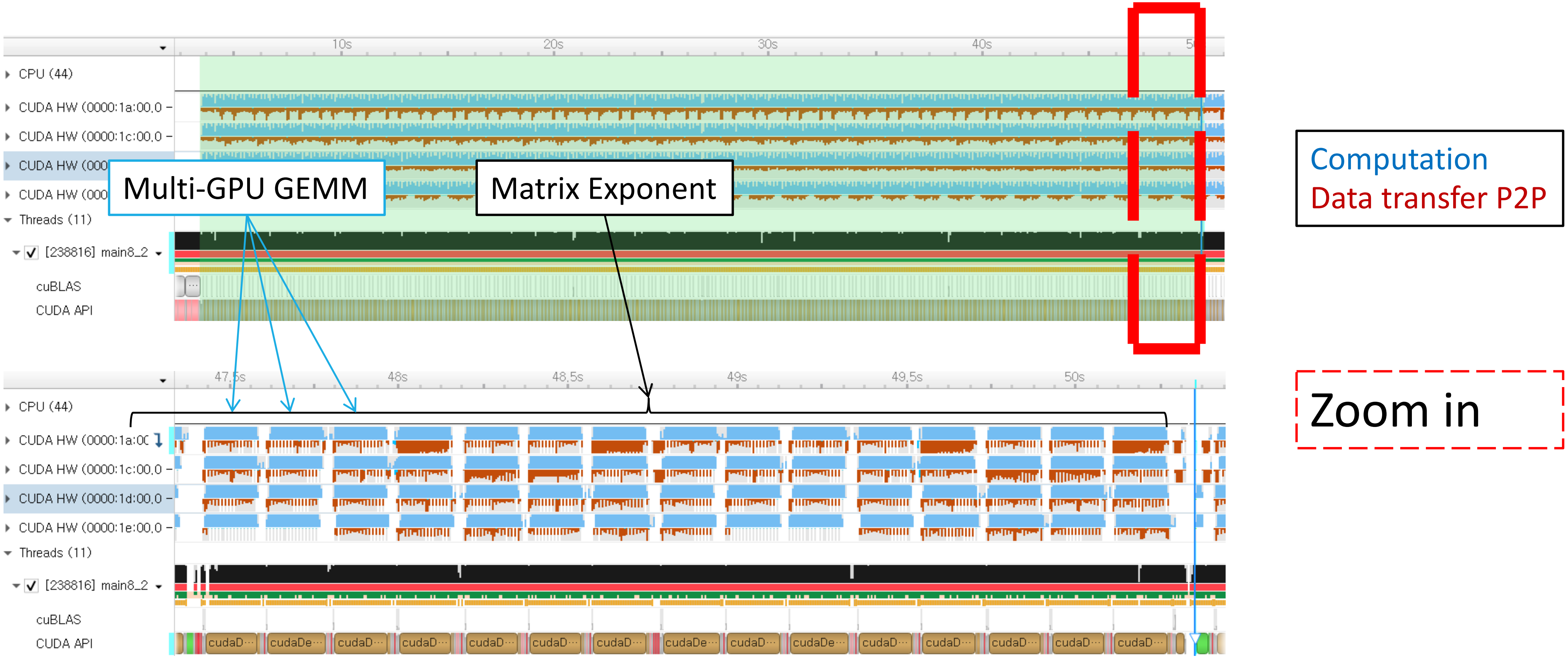
Computation time on nodes of supercomputer «cHARISMa»



One time iteration runtime (sec) with matrix A sized $N * N$

Choi Y. R., Nikolskiy V., Stegailov V. Tuning of a Matrix-Matrix Multiplication Algorithm for Several GPUs Connected by Fast Communication Links, in: *Parallel Computational Technologies: 16th International Conference, PCT 2022, Dubna, Russia, March 29–31, 2022, Revised Selected Papers*. Springer, 2022. Ch. 12. P. 158-171.

Profile of TDSE Algorithm on 4*V100 GPUs



Summary

- Developed and studied brute-force approach for 1D TDSE algorithm based on arbitrary matrix exponent with multi-GPU GEMM algorithm implementation with no assumptions on the structure of the Hamiltonian matrix and on the system symmetry
- High speed recalculation of the Hamiltonian exponent opens the possibility to study the processes that include the coupled electron-ion quantum dynamics
- Non-adiabatic (vibronic) energy transfer was modelled from moving nuclei to the single electron excitation of H_2^+