Acceleration of Ultrasound Neurostimulation Using Mixed-Precision Arithmetic

Jiri Jaros and Radek Duchon

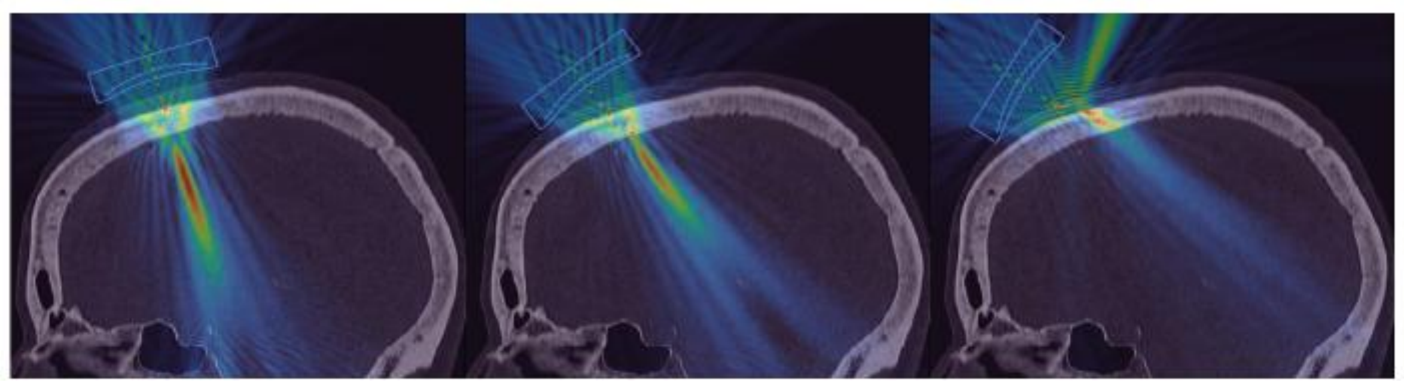
Brno University of Technology, Faculty of Information Technology, Brno, CZ

1 Introduction and Motivation

Brain disorders, encompassing both neurological and psychiatric conditions, impact one in every four individuals. Ultrasound neurostimulation, a technique that modulates the brain's electrical activity, has emerged as a significant secondary treatment option for cases resistant to pharmacological interventions.

The ideal solution would be a method capable of targeting specific brain regions with high anatomical accuracy, offering unlimited penetration depth, complete reversibility, and a minimal risk profile.

This objective is achievable through the application of a three-dimensional steerable ultrasound, directed by patient-specific stimulation plans. These plans are meticulously crafted through full-wave ultrasound propagation simulations. Nonetheless, the computational intensity required for calculating these plans poses a significant challenge, often reaching the memory capacities of contemporary graphics processing units (GPUs).



2 Modeling Ultrasound Wave Propagation in Brain

The propagation of ultrasound waves through the skull and within the brain is modeled by the governing wave equations outlined in the k-Wave toolbox (www.k-Wave.org).

These equations are resolved using a pseudospectral method that extensively employs Fourier transforms for the calculation of spatial gradients, alongside k-space correction for in-time integration. This approach is designed to minimize spatial and temporal resolution to the greatest extent possible.

$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{1}{\rho_0}\nabla p$$ momentum conservation

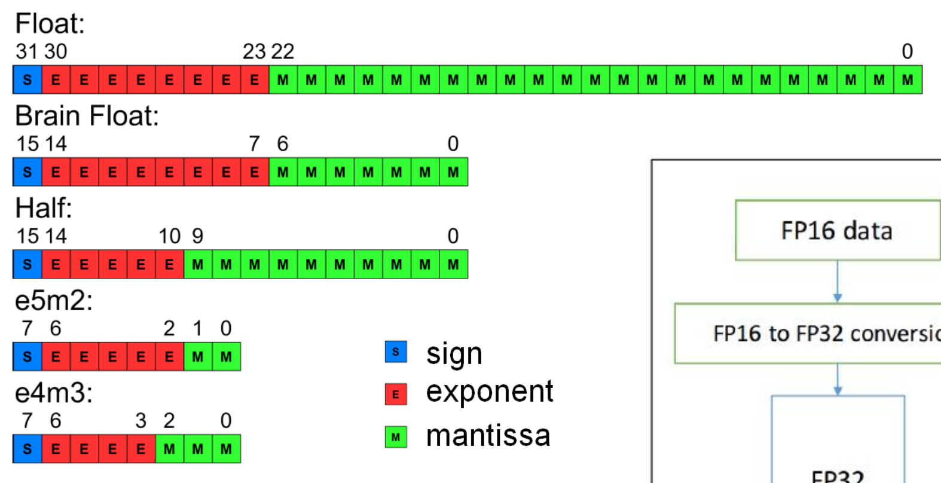
$$\frac{\partial \rho}{\partial t} = -\left(2\rho + \rho_0\right)\nabla\cdot\mathbf{u} - \mathbf{u}\cdot\nabla\rho_0$$ mass conservation

$$p = c_0^2\left(\rho + \mathbf{d}\cdot\nabla\rho_0 + \frac{B}{2A}\frac{\rho^2}{\rho_0} - \Pi\rho\right)$$ pressure-density relation

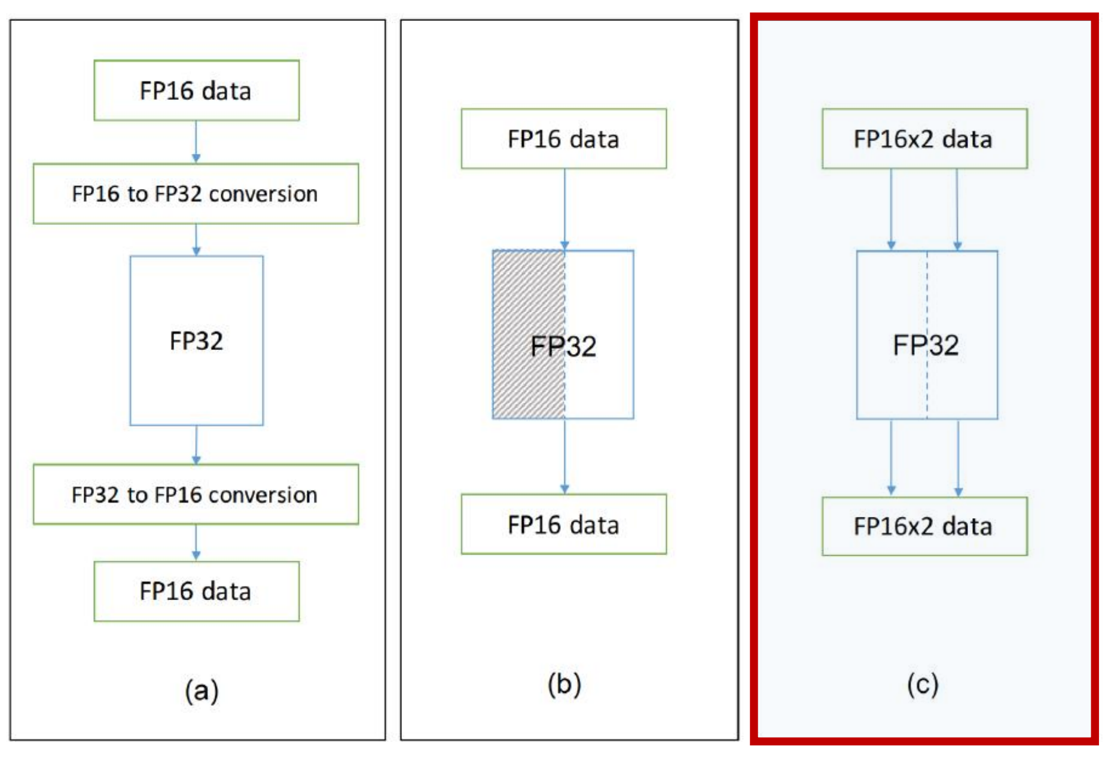
3 Reduced Precision Data Types

An innovative strategy showcased in this approach is the reduction of the simulation's memory requirements by utilizing reduced precision data types for storing specific quantities.

This not only diminishes the memory footprint but also accelerates execution by enabling the simultaneous performance of two operations in half precision. Additionally, an investigation into the numerical accuracy has been conducted to ensure the reliability of the simulations despite the reduced precision, highlighting a balance between computational efficiency and precision.



- Float
  - Maximum: ~3.39e+38
  - $\sum_{n=1}^{\infty} 1/n = \sim 15.404$
- Half
  - Maximum: 65504
  - $\sum_{n=1}^{\infty} 1/n = \sim 7.0859$
- Brain float
  - Maximum: ~3.39e+38
  - $\sum_{n=1}^{\infty} 1/n = \sim 5.0625$



6 Conclusions

By representing material properties and k-space operators more efficiently, we achieved up to 22% reduction in GPU memory usage, while accelerating calculations by 8.5% on Nvidia V100 GPU. This optimization introduced an error that reduced focal pressure by 0.5% without any focus movement, values that are clinically acceptable.

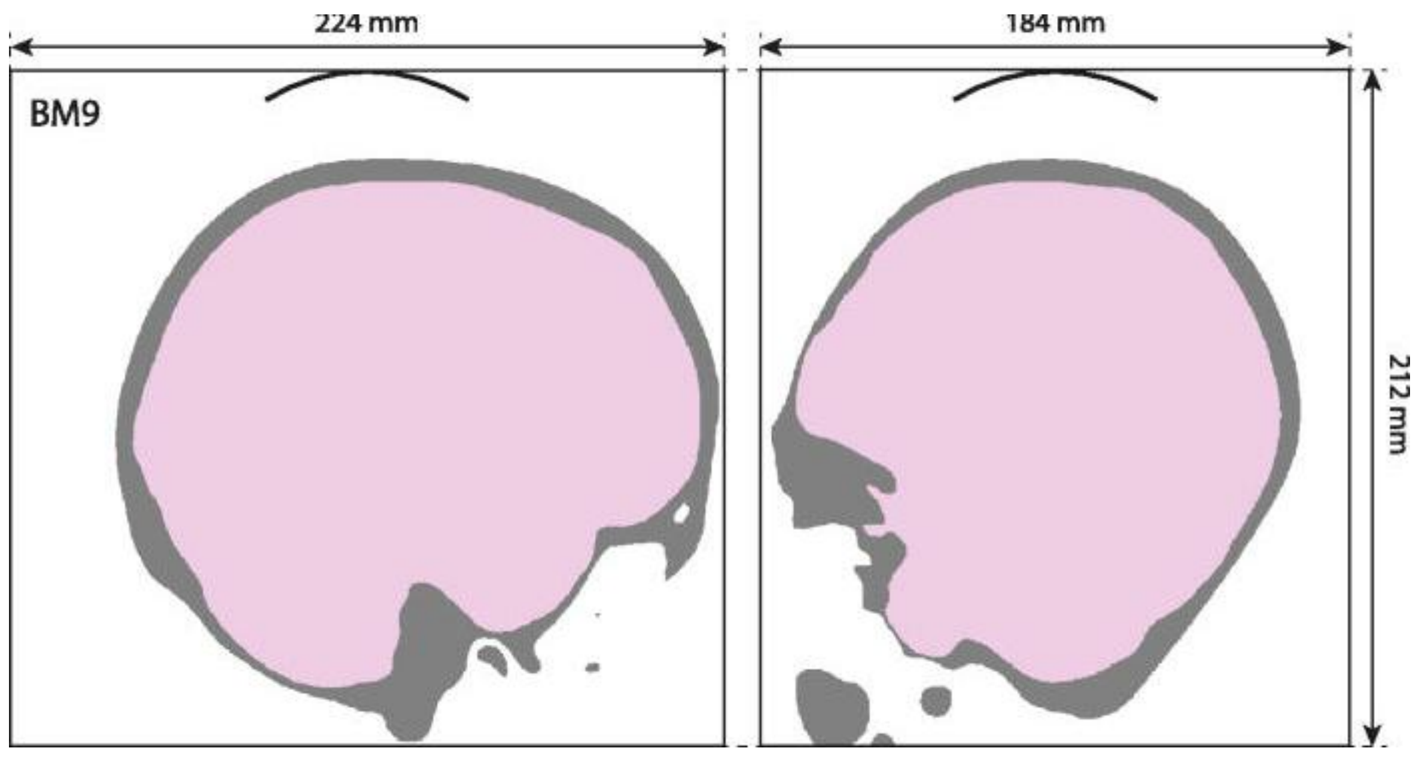
4 Achieved Memory Reduction

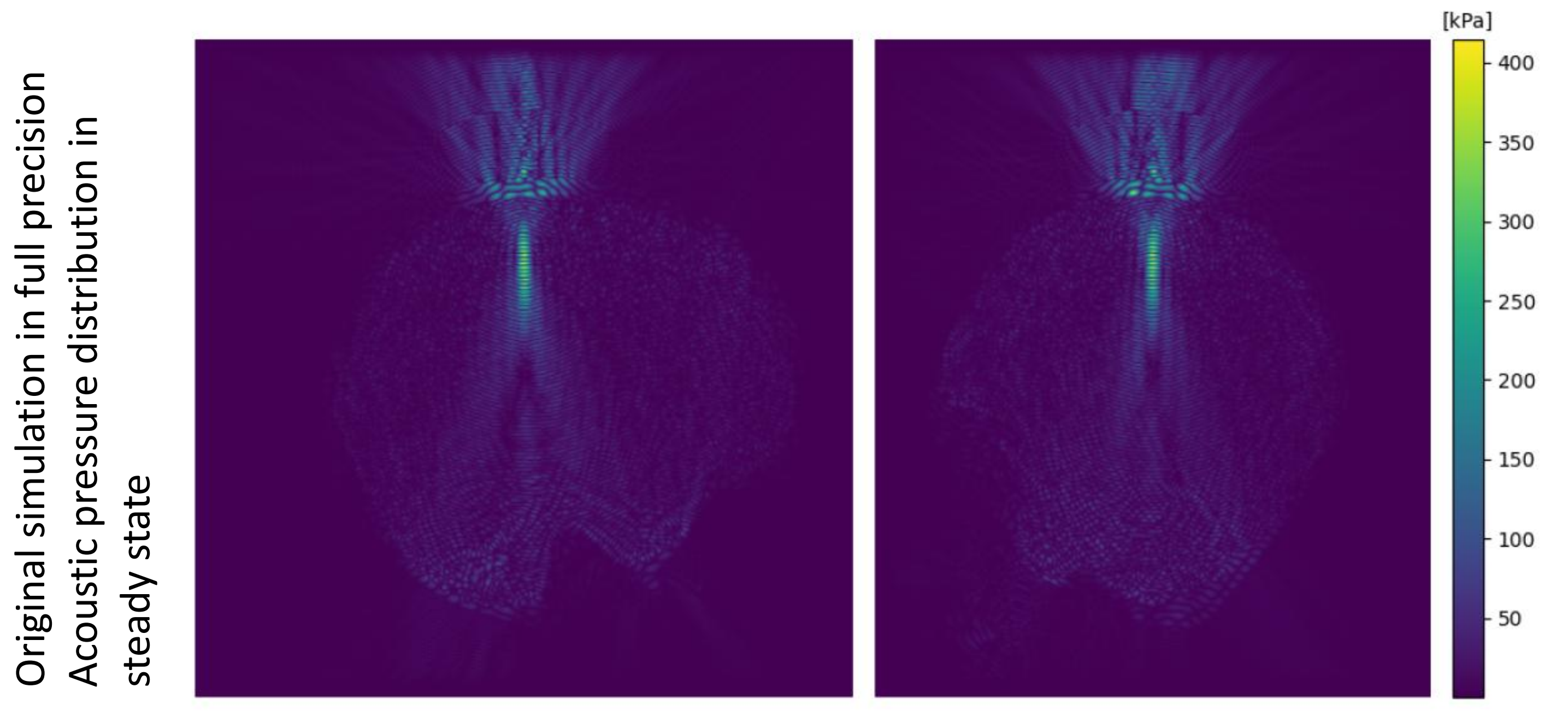
Due to the high dynamic range needed for Fourier transforms and the disparate units of measure used for acoustic pressure and acoustic particle velocity, these quantities and their gradients are excluded from reduction. However, material properties, with their 3-4 digit precision, are suitable for reduced data types. This study explores three levels of memory reduction, balancing computational efficiency with accuracy.

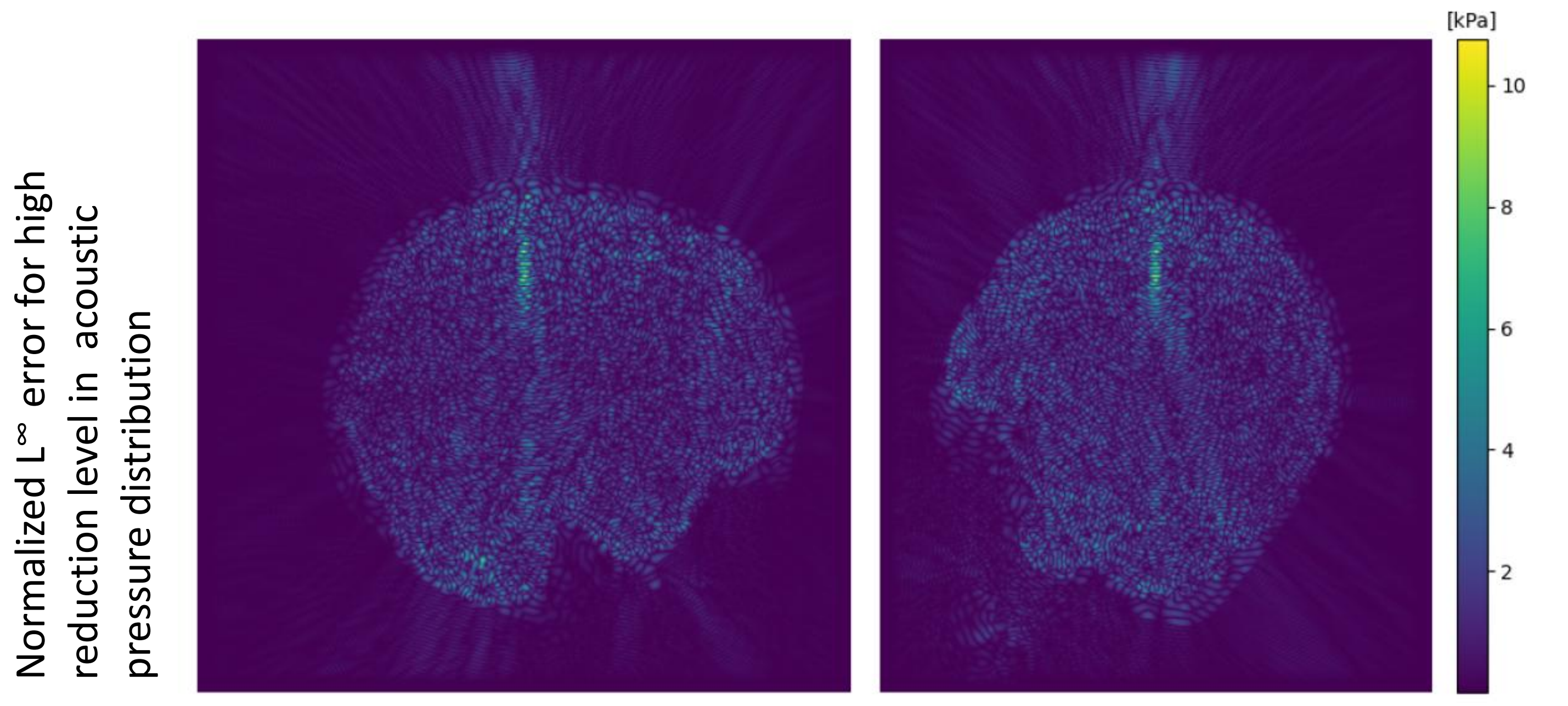
		Reduction Level [# matrices]		
Quantity	Data type	Low	Mid	High
Nonlinearity Coefficient	Half	1	1	1
Sound Speed	Half	1	1	1
Density	Half	1	4	4
Reference Density	Half	0	0	3
Initial Pressure	Brain Float	1	1	1
Pressure Source	Brain Float	1	1	1
Absorption Tau and Eta	Brain Float	2	2	2
Absorption Nabla	Brain Float	1	1	1
Kappa Operator	Half	0	0	0.5
Source Kappa Operator	Half	0.5	0.5	0.5
Reduced Matrices		8.5	11.5	15
Memory Reduction		10.5%	15.7%	21.8%
Execution Time Reduction		3.6%	6.9%	8.5%

5 Evaluation of Simulation Accuracy

A whole-skull ultrasound neurostimulation simulation was performed using a 512x512x432 grid over 12,000 time steps. The comparison reveals that the maximum acoustic pressure in the original simulation and the relative absolute error from the high memory-conserved simulation showed a maximum error of 2.9%. The focal amplitude difference remained below 0.5% without any focal movement.




Original simulation in full precision
Acoustic pressure distribution in steady state


Normalized L$^\infty$ error for high reduction level in acoustic pressure distribution

Error Metric	Accuracy Loss / Reduction Level Low	Mid	High
L^2 Error Norm	0.092%	0.731%	6.98%
L^∞ Error Norm	0.106%	0.542%	2.941%
Focal Amplitude Difference	0.070%	0.132%	0.530%
Focus Movement	0	0	0

This project has received funding from the European Unions Horizon Europe research and innovation programme under grant agreement No 101071008. This work was also supported by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90254), and by Brno University of Technology under project FIT-S-23-8141.