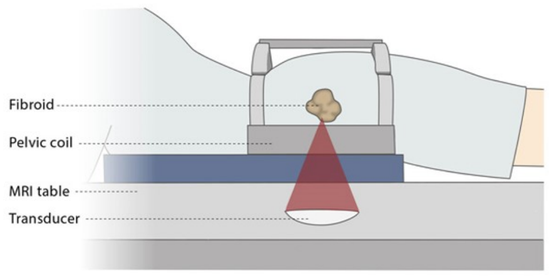
# Running Large-Scale Ultrasound Simulations on Intel Xeon Phi (KNC) accelerators

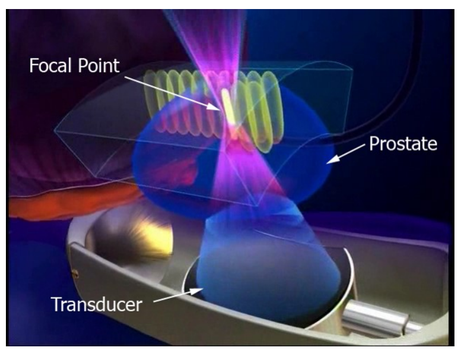
Filip Vaverka[1], Bradley E. Treeby[2] and Jiri Jaros[1]

[1]Faculty of Information Technology, Brno University of Technology, Centre of Excellence IT4Innovations, CZ
[2]Department of Medial Physics and Biomedical Engineering, University College London, UK

## Overview

High-intensity focused ultrasound (HIFU) is an emerging non-invasive cancer therapy that uses tightly focused ultrasound waves to destroy tissue cells through localised heating. The treatment planning goal is to select the best transducer position and transmit parameters to accurately target the tumour. The path of the ultrasound waves can be predicted by solving governing acoustic equations based on mass, momentum and energy conservation. However, this is a computationally difficult problem because the domain size is very large compared to the acoustic wavelength.




## Nonlinear Ultrasound Wave Propagation in Tissue

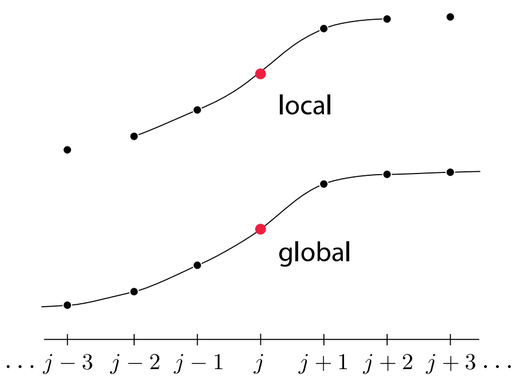
The governing equations account for the nonlinear propagation of ultrasound waves in tissue, which is a heterogeneous and absorbing medium. Accurate acoustic absorption is critical for predicting ultrasound dose under different conditions. The required acoustic equations can be written as:

$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{1}{\rho_0}\nabla p + \mathbf{S}_F \qquad \text{(momentum conservation)}$$

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

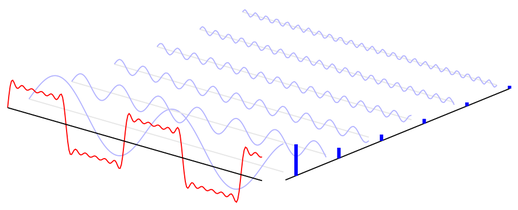
$$p = c_0^2\left(\rho + \mathbf{d} \cdot \nabla \rho_0 + \frac{B}{2A}\frac{\rho^2}{\rho_0} - L\rho\right) \qquad \text{(pressure-density relation)}$$

These equations are discretized using the k-space pseudo-spectral method and solved iteratively. This reduces the number of required grid points per wavelength by an order of magnitude compared to finite element or finite difference methods. For uniform Cartesian grids, the gradients can be calculated using the fast Fourier transform (FFT).




$$\frac{\partial f}{\partial x} = \frac{\frac{1}{12}f^{j-2} - \frac{2}{3}f^{j-1} + \frac{2}{3}f^{j+1} - \frac{1}{12}f^{j+2}}{\Delta x}$$

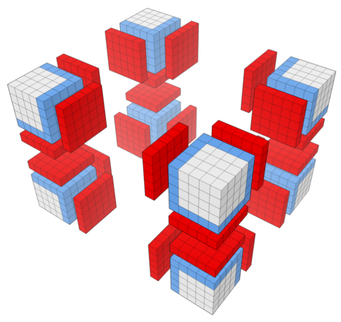
$$\frac{\partial f}{\partial x} = \mathscr{F}^{-1}\{ik_x\mathscr{F}\{f\}\}$$

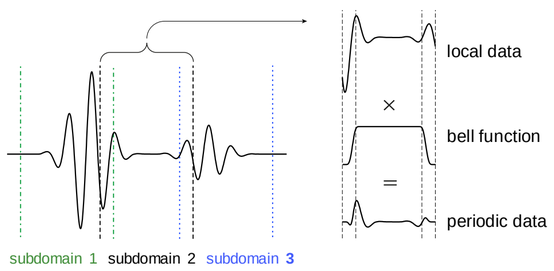


## Local Fourier Basis Decomposition

Local Fourier basis decomposition reduces the communication burden by partitioning the domain into a grid of local subdomains where the gradients are calculated locally and the global communication is replaced by the nearest-neighbor overlap exchange. The gradient calculation on each sub-domain then reads as follows, where $b$ is a bell function smoothing the subdomain interface:

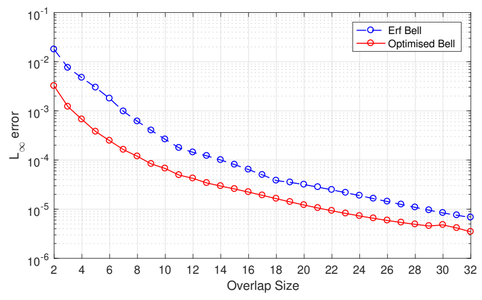
$$\frac{\partial p}{\partial x} = \mathscr{F}^{-1}\{ik\mathscr{F}\{b \circ p\}\}$$

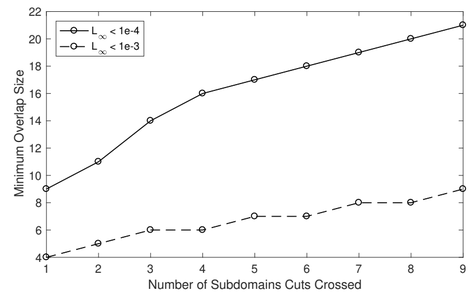





## Local Fourier Basis Accuracy

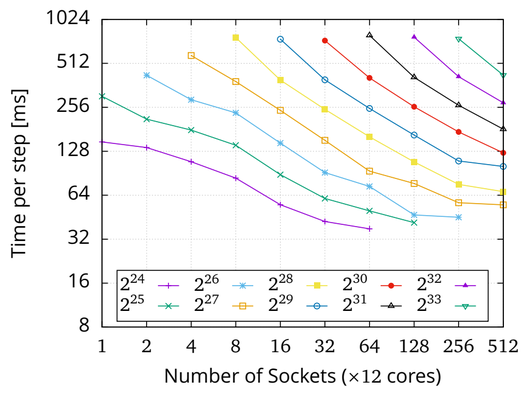
Since the gradient is not calculated on the whole data, numeric error is introduced. Its level can be tuned by the shape of the bell function and thickness of the overlap region. An overlap size of 16 grid points is then sufficient to maintain the $L_\infty$ error below 0.01% even after the wave has crossed 8 subdomain boundaries. For 3D decompositions, this corresponds to $9^3 = 729$ local subdomains.

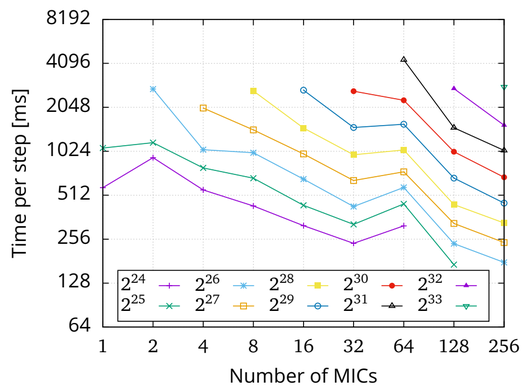



## Simulation Scaling

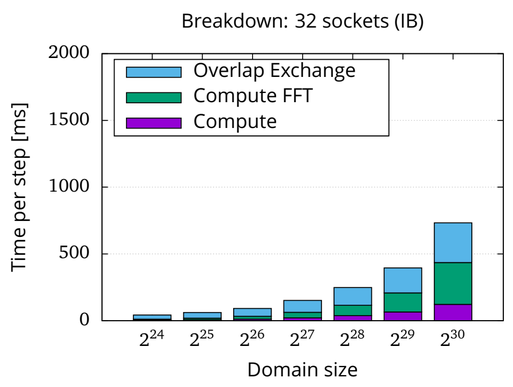
The proposed code running on Intel Xeon Phi accelerators (MICs) with **TCP** Intel MPI backend (as **DAPL** is not stable enough) achieves scaling comparable to the CPU-only version running over Infiniband (albeit with significant slowdown). The average speedup when doubling the number of sockets or accelerators is around **60%** (for 16-point deep overlaps).

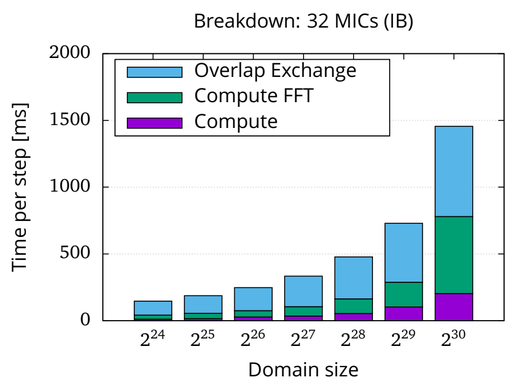


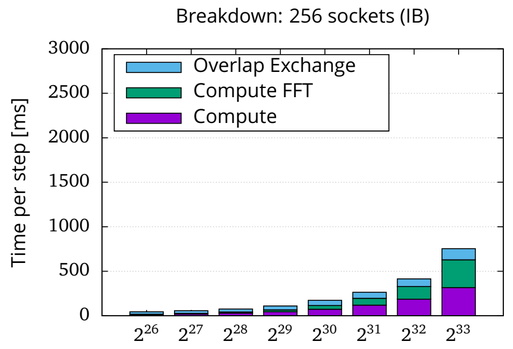



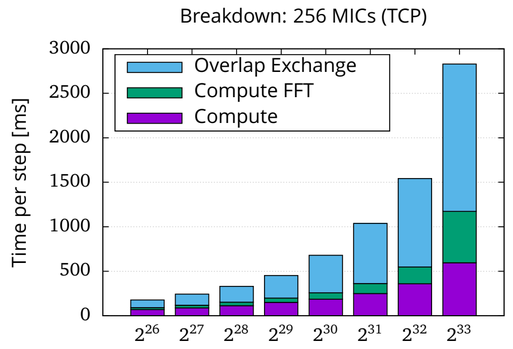

## Decomposition Overhead

Overall, the simulation is about **2× slower** than the CPU version when running on accelerators connected via Infiniband, or **4× slower** when Ethernet has to be used.




## Intel Xeon Phi Support

The Intel Xeon Phi accelerators on Salomon suffer from multiple more or less significant issues which are rather difficult to workaround.

- Parallel I/O on LustreFS (Intel MPI-IO / HDF5) is no longer supported.
- Inifiniband Intel MPI backend (DAPL) is unstable/unusable when more than 32 accelerators/processes are involved.
- Performance of Intel MKL on Xeon Phi is rather limited.

k-Wave
A MATLAB toolbox for the time-domain simulation of acoustic wave fields

IT4Innovations national supercomputing center

Biomedical Ultrasound Group

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project "IT4Innovations excellence in science - LQ1602" and by the IT4Innovations infrastructure which is supported from the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Supercomputing Center - LM2015070". This work was also supported by the Engineering and Physical Sciences Research Council, UK, grant number EP/L020262/1. This project has received funding from the European Union's Horizon 2020 research and innovation programme H2020 ICT 2016-2017 under grant agreement No 732411 and is an initiative of the Photonics Public Private Partnership.