

Russian supercomputing days '2017



OFFICIAL WEBSITE

Архитектура специализированных GPU-кластеров для решения обратных задач 3D низкочастотной ультразвуковой томографии.

Гончарский А.В., Серёжников С.Ю. (докладчик)

Исследования проведены при поддержке гранта РНФ № 17-11-01065.

Московский государственный университет им. М. В. Ломоносова, НИВЦ

Ультразвуковая томография в медицине



- **Рентген**

Высокая доза облучения при
регулярных обследованиях

- **МРТ**

Высокая стоимость
и сложность оборудования

- **УЗИ**

Недостаточное качество
изображения для ранней
диагностики

- Более 40 тысяч женщин ежегодно заболевают раком молочной железы. Доля лиц с поздними стадиями заболевания среди первичных больных превышает 40%.
- Каждое 4-е онкологическое заболевание – рак груди.
- Ультразвуковые томографы высокого разрешения позволяют осуществлять раннюю диагностику рака.
- В настоящее время не существует серийно выпускаемых ультразвуковых томографов.
- В США и Европе ведутся интенсивные работы по разработке программного обеспечения и макетов установок.

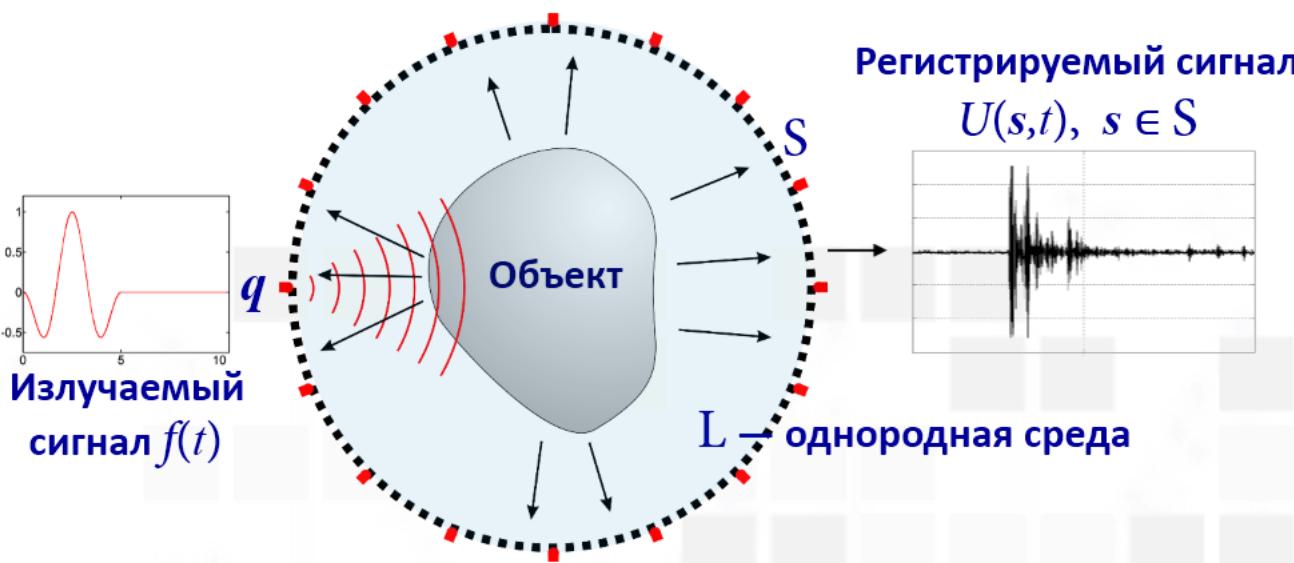
Методы ультразвуковой диагностики

3

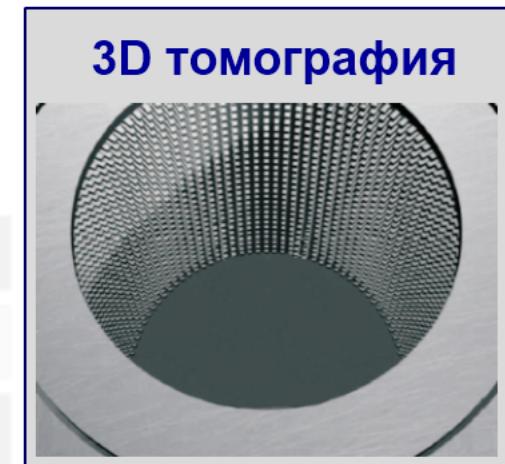
	Количество трансдьюсеров	Сложность восстановления	Качество изображения
УЗИ	100(2D) – 10000(3D)	0.01с	--
Линейная томография	1000 (2D)	1с	-
Приближённые волновые модели	1000 (2D)	10 – 100с	+
Синтезированная апертура	100 000+ (3D)	10 – 100с	+
➤ Коэффициентная обратная задача	10000 (3D)	10 000с	++
Distorted Born Iterative (функция Грина)	10000 (3D)	1 000 000с	++

Постановка задачи волновой томографии

4



3D томография



- Волновое уравнение

$c(\mathbf{r})u_{tt}(\mathbf{r},t) + a(\mathbf{r})u_t(\mathbf{r},t) - \Delta u(\mathbf{r},t) = \delta(\mathbf{r}-\mathbf{q}) \cdot f(t), \quad \mathbf{q} - \text{положение источников},$

$u(\mathbf{r},0)=0; \quad u_t(\mathbf{r},0)=0; \quad \partial_n u|_{\text{ST}}=p(\mathbf{r},t) \quad - \text{начальные и граничные условия (известны)},$

$u(s,t)=U(s,t) \quad - \text{поле, измеренное на приёмниках (известно на части границы)},$

$c(\mathbf{r})=c_0, \quad a(\mathbf{r})=a_0 \quad \text{в однородной среде} \quad \mathbf{r} \in L.$

- Требуется найти $c(\mathbf{r}), a(\mathbf{r})$. Нелинейная обратная задача.

Постановка задачи волновой томографии

- Задача – минимизация функционала невязки

$$\Phi(u(\textcolor{red}{c}, \textcolor{red}{a})) = \| u(s, t) - \textcolor{green}{U}(s, t) \|$$

Основная задача

$$t=0 \dots T$$

$$c(\mathbf{r})u_{tt}(\mathbf{r}, t) + a(\mathbf{r})u_t(\mathbf{r}, t) - \Delta u(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{q}) \cdot \textcolor{green}{f}(t)$$

$$u(\mathbf{r}, 0) = 0; \quad u_t(\mathbf{r}, 0) = 0;$$

$$\partial_n u|_{\text{ST}} = \textcolor{green}{p}(\mathbf{r}, t)$$

- Градиент функционала имеет вид:

$$\Phi'_c(u) = \int_0^T u_t(\mathbf{r}, t) w_t(\mathbf{r}, t) dt \quad \Phi'_a(u) = \int_0^T u(\mathbf{r}, t) w_t(\mathbf{r}, t) dt$$

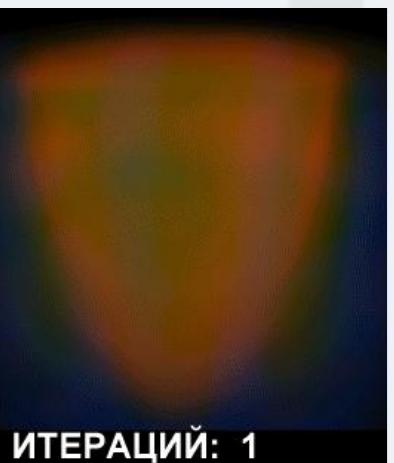
- Градиент суммируется по всем источникам q .

- Итерационный метод минимизации функционала

$$c(0) = \text{const}; \quad a(0) = 0;$$

50...100 итераций

$$\{c, a\}_{(n+1)} = \{c, a\}_{(n)} + \lambda \{\Phi'_c(u(c_{(n)}, a_{(n)}), \Phi'_a(u(c_{(n)}, a_{(n)}))\}$$



ИТЕРАЦИЙ: 1

Численный алгоритм

- Явная разностная схема для расчета распространения звуковой волны последовательно по времени
- Аппроксимации частных производных 2-го порядка точности
- Дискретный Лапласиан размерности 3×3×3 и выше

6

$$c_{ijl} \frac{u_{ijl}^{k+1} - 2u_{ijl}^k + u_{ijl}^{k-1}}{\tau^2} + a_{ijl} \frac{u_{ijl}^{k+1} - u_{ijl}^{k-1}}{\tau} - \frac{\Delta u_{ijl}^k}{h^2} = 0$$

$$\Delta u_{i_0 j_0 l_0}^k = \sum_{i=i_0-1}^{i_0+1} \sum_{j=j_0-1}^{j_0+1} \sum_{l=l_0-1}^{l_0+1} b_{ijl} u_{ijl}^k$$

- Градиент функционала невязки вычисляется по формуле

$$(\Phi'_c)_{ijl} = \sum_{k=2}^{M-2} \frac{u_{ijl}^{k+1} - u_{ijl}^k}{\tau} \frac{w_{ijl}^{k+1} - w_{ijl}^k}{\tau} \tau$$

$$(\Phi'_a)_{ijl} = \sum_{k=2}^{M-2} u_{ijl}^k \frac{w_{ijl}^{k+1} - w_{ijl}^{k-1}}{\tau} \tau$$

- Используется 3 слоя по времени

Численный алгоритм – GPU

- Z-marching
- Размерность задачи $\sim 400^3$
- Кол-во параллельных потоков ~ 10000

• Расчётная формула

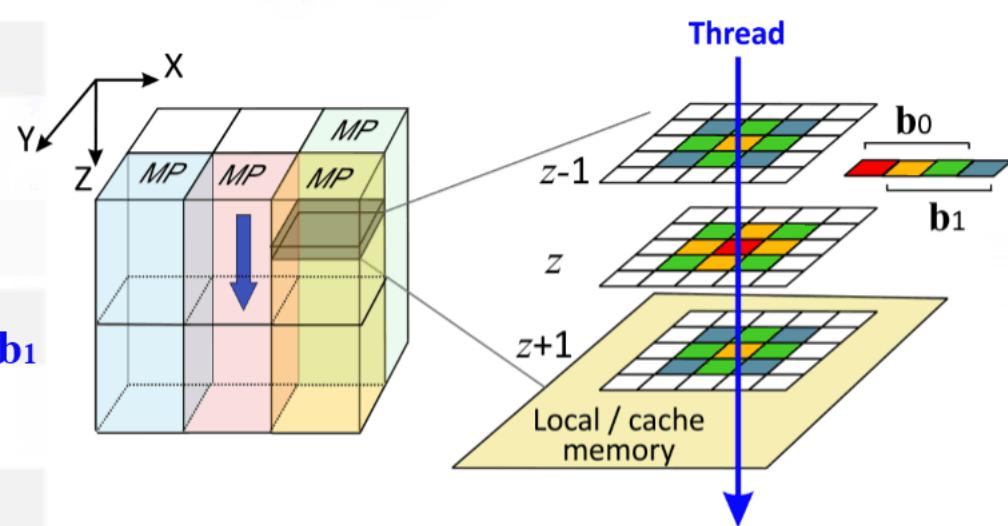
$$u(r, t+1) = K_1 u(r, t-1) + K_2 [u(r, t) \cdot b]$$
$$[u(r, t) \cdot b] = u(z) \cdot b_0 + (u(z-1) + u(z+1)) \cdot b_1$$

• Memory-intensive

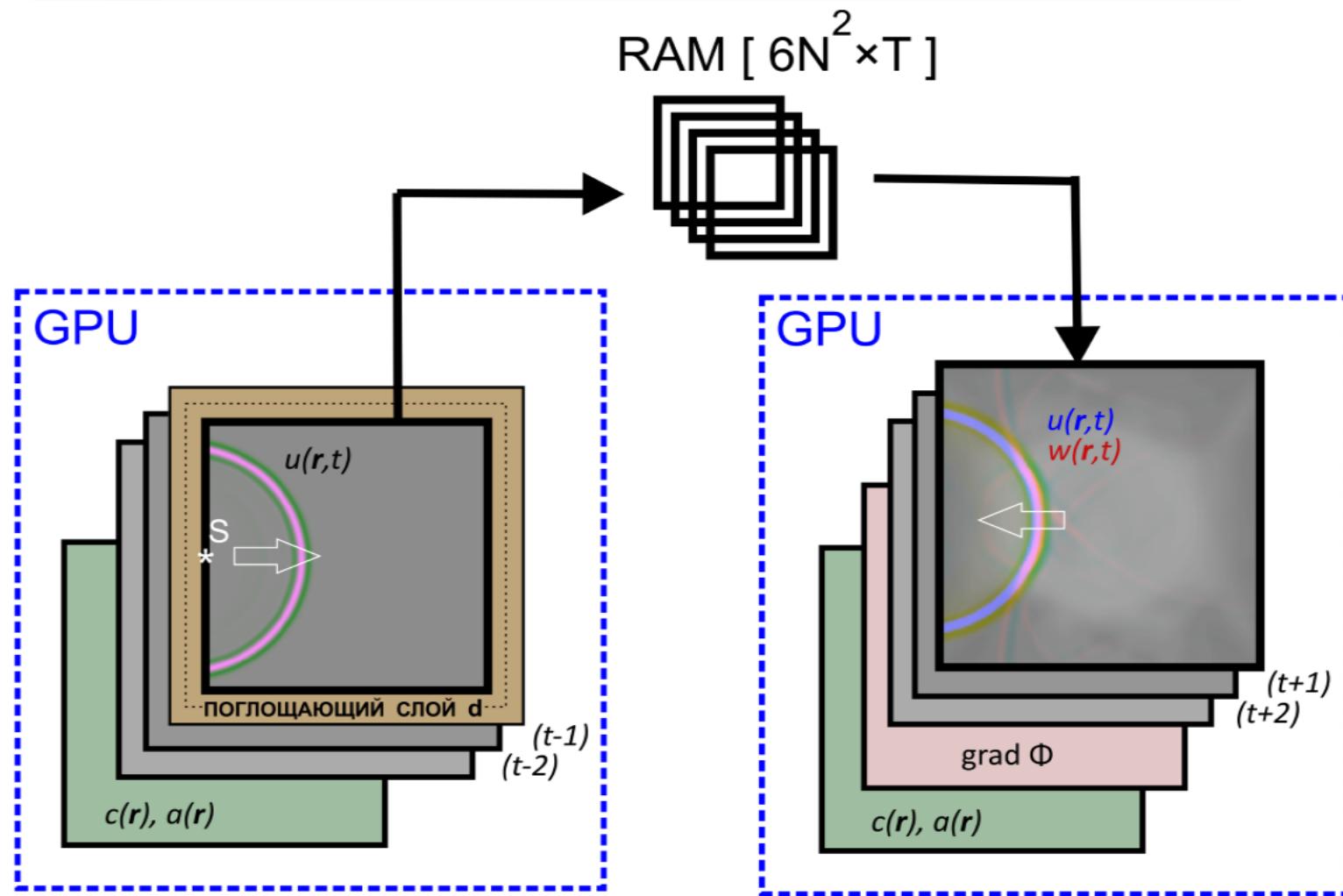
• Специфика задачи — искомый градиент является интегралом по времени

$$\Phi'_c(u) = \int u_t(r, t) w_t(r, t) dt$$

• Достаточно хранить только 3 слоя по времени



Метод расчёта

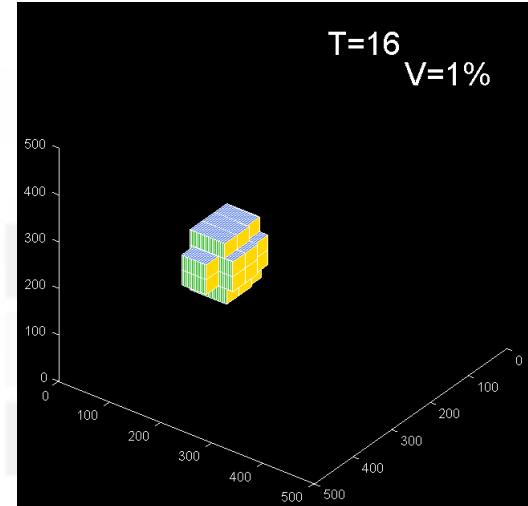


Оптимизация

Устройства: NVidia GeForce GTX 660, GTX TITAN, Tesla X2070, Tesla K40s
Интерфейс OpenCL

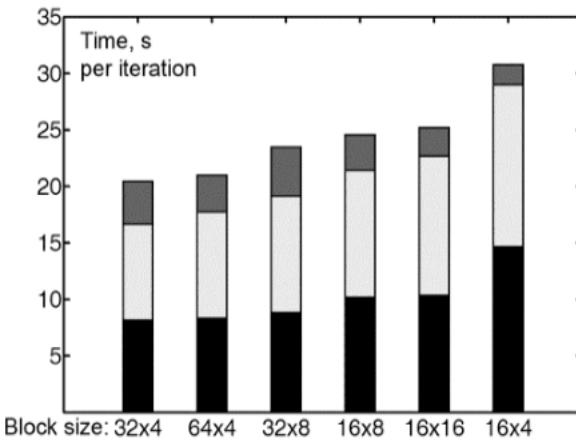
1) Выбор представления данных

- Акустическое давление u, w — 32 бит (float)
- Коэффициенты уравнения c, a — 2×16 бит (half, read-only image3d)
- Измеряемые данные — 16 бит (short)
- Градиент — 2×16 бит (short)



2) Специфика — низкий контраст

- Скорость звука близка к постоянной
- Динамическое изменение объёма (~65%)



3) Распределение данных по устройствам GPU

- Оптимальный размер блоков 32×4 [$\times 32$]
- Использование локальной памяти (GTX)
- Автоматическое кэширование (Tesla)
- Использование constant cache

Системные требования

- На каждой итерации для каждого источника вычисляется $u(r,t), w(r,t)$; $r: N^3 \approx 50$ млн. точек, $t: \approx 1000$ шагов
- Общая вычислительная сложность для 100 итераций, 50 источников — 10^{15}
- Приемлемое на практике время расчёта — **1 час**
- Требуемый объём памяти $\sim N^3$
- Количество операций $\sim N^4$
- Требуемая производительность GPU-кластера **10 – 20 ТБ/с (1000× CPU)**

Размерность задачи (N)	256	288	320	384	416	448	512
Память GPU, GB	0.8	1.2	1.5	2.2	2.8	3.5	5.0
Системная память, GB	1.5	2	3	5	6	8	12
Время (1 итерация), сек	11	16	23	48	63	92	160

Low-end GPU

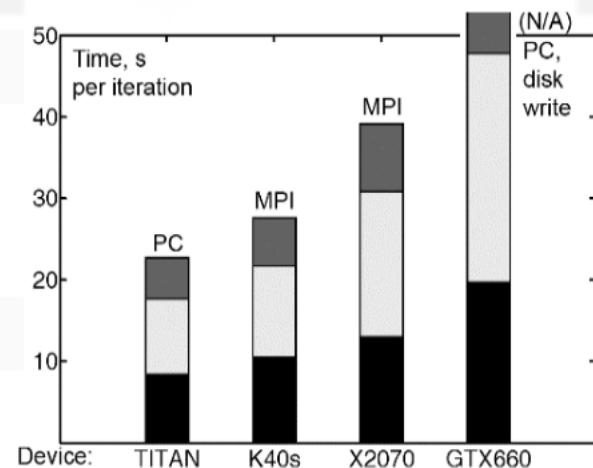
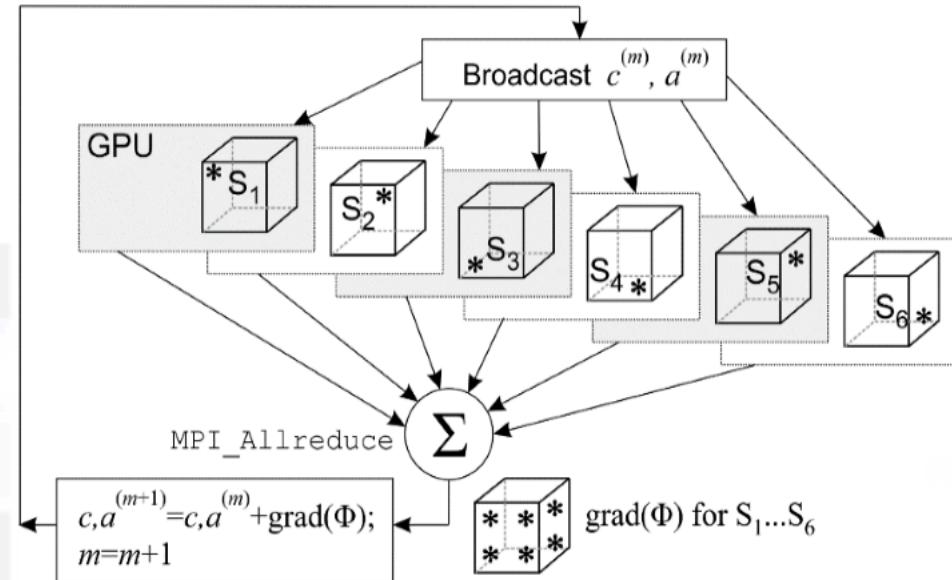
GTX TITAN

1x HBM-class
2x GPU

2x HBM
4x GPU

Распараллеливание вычислений

- Каждый GPU производит вычисления для одного источника
- После каждой итерации результаты для всех источников суммируются.
- Тестирование проводилось на СК “Ломоносов” и “Ломоносов-2” СКЦ МГУ – сеть QDR Infiniband **40ГБит/с**
- Интерфейс MPI
- Время передачи данных **0.3 сек. (0.5%)**
- Минимальная пропускная способность сети – **200 МБайт/с**



Архитектура GPU-кластера

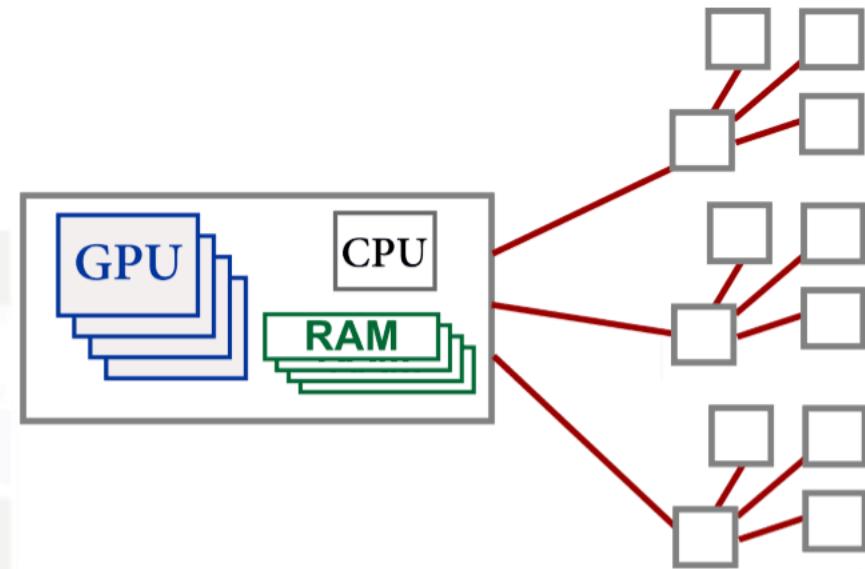
- На каждый узел:

Кол-во GPU → max (4)

память GPU 3 – 6 GB

RAM — 32 GB

Кол-во CPU → min (1)



- Сеть ориентирована на суммирование и рассылку

Пропускная способность > 200 MB/s

- Общее кол-во GPU ~40, узлов 10 .. 20
- Потребляемая мощность <15 kW
- Габариты — 1 стойка

Основные результаты

- Предложены эффективные методы и алгоритмы решения 3D задач волновой томографии как коэффициентных обратных задач через прямое вычисление градиента функционала невязки.
- 3D нелинейная обратная задача с числом неизвестных порядка 100 млн. имеет вычислительную сложность порядка 10^{15} .
- Предложена структура GPU-кластера, который может входить в состав ультразвукового томографического комплекса.

Работа выполнена при поддержке гранта РНФ № 17-11-01065

Наш сайт : <http://inverseproblems.ru>

Спасибо за внимание!