

Дослідження методів оптимізації та паралелізації математичних обчислень для графічного процесора (GPU) у бібліотеці скінченних елементів MoFEM

Кваліфікаційна робота

Керівник кваліфікаційної роботи

канд. фіз-мат. наук, доцент

Бублик В.В.

Виконав студент Шевченко Б. Ю.



NATIONAL UNIVERSITY OF
KYIV-MOHYLA ACADEMY



UNIVERSITY
of
GLASGOW

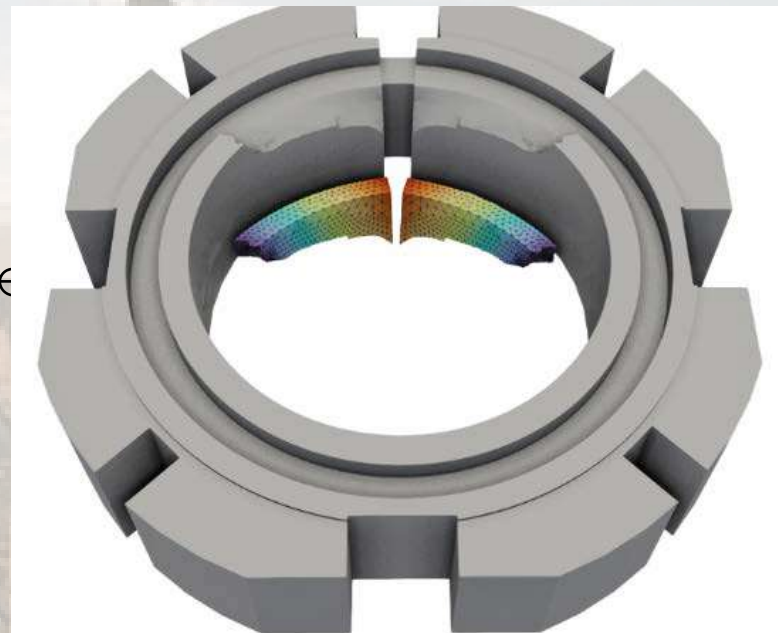
Актуальність проблеми

Чому це важливо?

- Безпека атомних електростанцій
- Радіаційне опромінення графітових блоків
- Це призводить до створення тріщин
- Точне моделювання для прогнозування руйнувань

Проблеми:

- Мільйони ступенів свободи(змінних)
- Величезне навантаження на обчислювальні ре



Метод скінченних елементів (МСЕ)

3/12

Наближене розв'язування диференціальних рівняння з частинними похідними

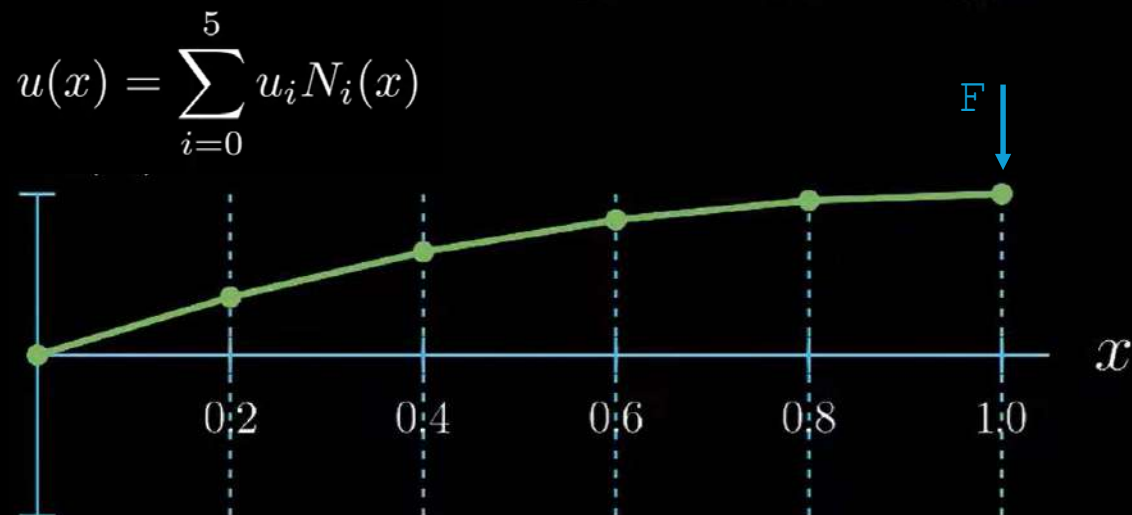
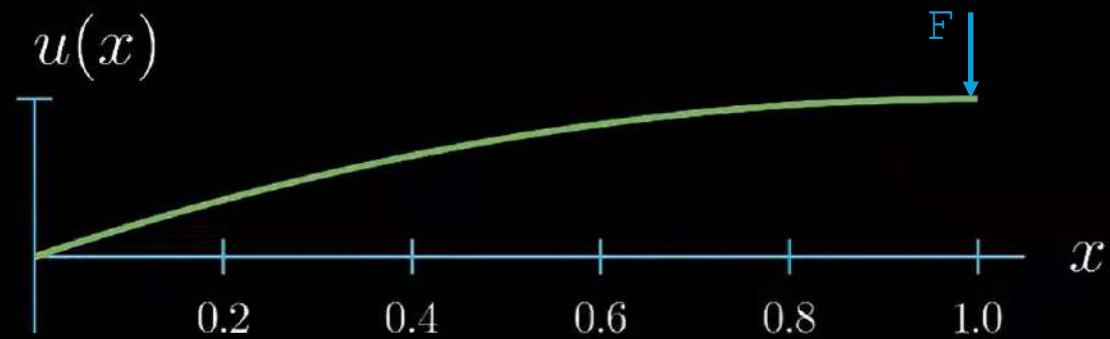
1. Розбиваємо функцію на багато лінійних ділянки
2. Кожна ділянка – це скінченний елемент
3. Будуємо лінійні рівняння для кожного елемента

Одержали:

Матриця жорсткості K та вектор навантажень F

Розв'язуємо рівняння $Ku = F$

$$\begin{bmatrix} 10 & -5 & 0 & 0 & 0 \\ -5 & 10 & -5 & 0 & 0 \\ 0 & -5 & 10 & -5 & 0 \\ 0 & 0 & -5 & 10 & -5 \\ 0 & 0 & 0 & -5 & 5 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = -\bar{f} \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.1 \end{bmatrix}$$



Змішані скінченні елементи

Чому звичайні елементи не завжди достатні?

- При моделюванні тріщин потрібна **висока точність напружень**
- Якщо переміщення апроксимуються поліномами степеня p , то напруження апроксимуються поліномами степеня $p - 1$

Змішані елементи:

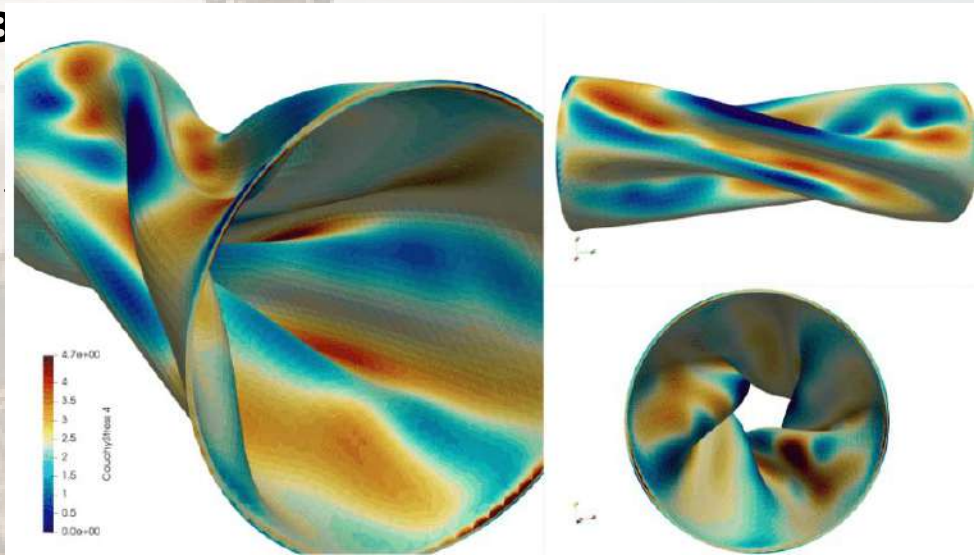
Шукаємо **одночасно** переміщення (u) і напруження (σ)

Шукаємо розв'язок в **добутку функціональних просторів**: $(u, \sigma) \in U \times \Sigma$
 $\in U \times \Sigma$

Можна використовувати різні простори з різними в

Звичайно: $K \cdot u = F$

Змішано:
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \sigma \\ u \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$



Гібридизація

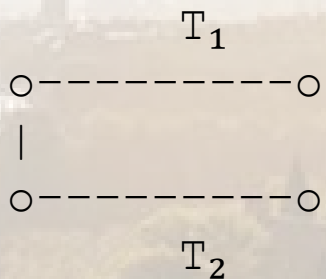
5/12

- Звичайна: 1M рівнянь
- Змішана: 2M рівнянь → в 8 разів повільніше!
- А мільйон – це лиш квіточки

Ключова ідея гібридизації:

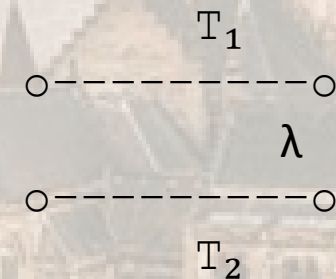
1. Замість: u, σ неперервні по всій області Ω
Робимо: u, σ можуть бути розривними на границях елементів

Вузол на границі = спільний для двох елементів



| ← цей вузол належить і T_1 , і T_2

Кожен елемент має свої локальні вузли



← λ на

Доповнення Шура

6/12

Що це таке?

Спосіб перетворення великої системи лінійних рівнянь у меншу

A (internal)	B ^T (coupling)	[u, σ] internal	=	f ₁
B (coupling)	θ (or small)	λ boundary		f ₂

Після перетворення: $S \cdot x_2 = g$
де $S = D - C \cdot A^{-1} \cdot B$ (доповнення Шура)

Результат: $S \cdot \lambda = f_2 - C \cdot A^{-1} \cdot f_1$

• Система зменшиться на порядки

D

Метод Крилова і GMRES

7/12

Навіщо потрібен спеціальний метод?

- Система все ще **дуже велика** (сотні тисяч рівнянь) – прямо розв'язати не вийде

GMRES (Generalized Minimal Residual Estimate Solver):

- **Поступово наближається** до точного розв'язку
- Шукає розв'язок в просторі **множень степенів матриці на початковий вектор**

Дослідження

8/12

Змішані гібридизовані скінченні елементи потребують дуже великих обчислень!
Найбільш ресурсоемними є обчислення задач лінійної алгебри.

Мета :

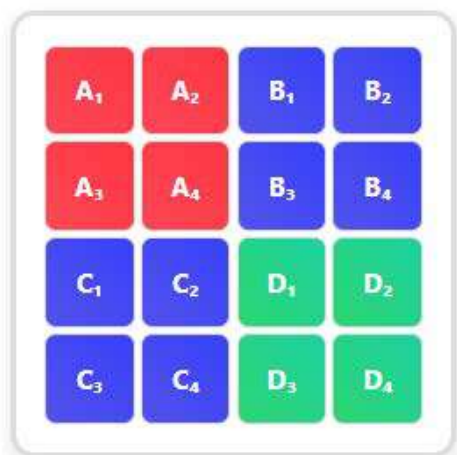
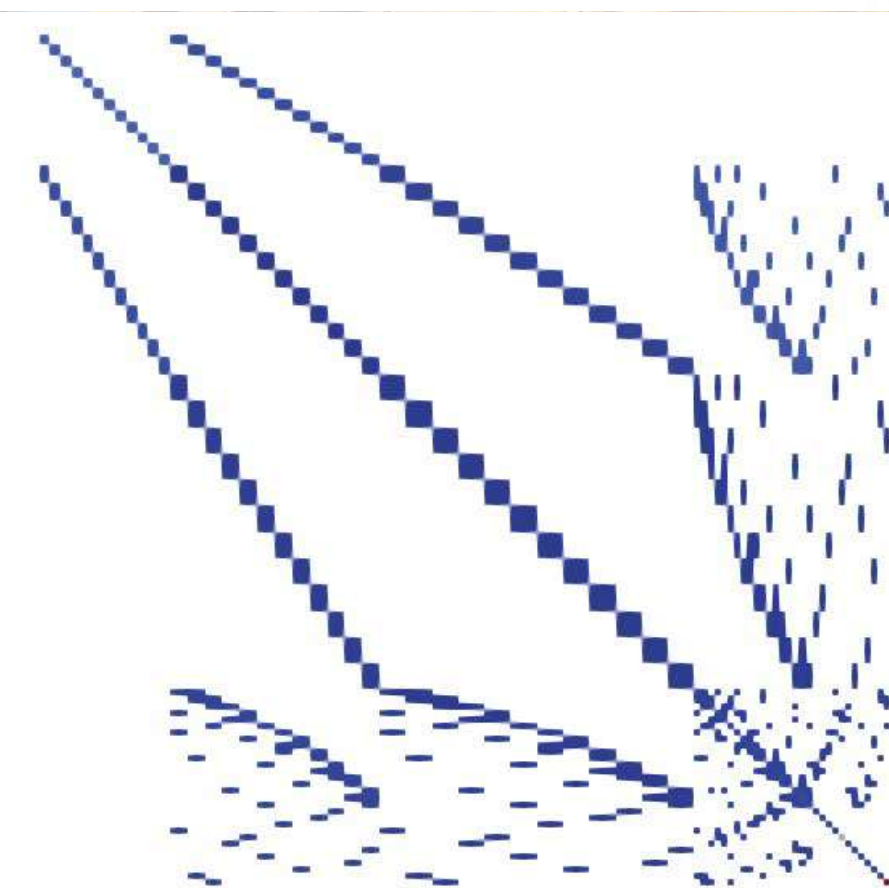
Підвищити продуктивність $MoFEM$ для моделювання тріщин через GPU-прискорення

Загальна архітектура – Матриця

9/12

Матриця є дуже розрідженою

- Зберігання відбувається в **блочному форматі BSE**
- **Густі блоки зберігаються в індексованій базі даних**
- Проблема змішує 3 функціональні простори:
 L_2 , L_2 , $Hdiv$



- A₀₀ - внутрішні блоки
- B, C - зв'язуючі блоки
- D - граничні блоки

Поблокове множення для методу Крилова

- Дані розділені між багатьма серверами
- При збірці матриці асинхронно передаються на локальний GPU
- Матриця знаходяться не у віртуальній пам'яті, а напряму у фізичній

Розріджена матриця

A_1	0	B_1	0	0
0	A_2	0	B_2	0
C_1	0	A_3	0	B_3
0	C_2	0	A_4	0
0	0	C_3	0	A_5

×

Вектор x

x_1
x_2
x_3
x_4
x_5

→

$$r_1 = A_1 \times x_1$$

$$r_2 = B_1 \times x_3$$

$$r_3 = A_2 \times x_2$$

$$r_4 = B_2 \times x_4$$

$$r_5 = C_1 \times x_1$$

$$r_6 = A_3 \times x_3$$

$$r_7 = B_3 \times x_5$$

...

Фінальний вектор y

y_1

y_2

y_3

y_4

y_5

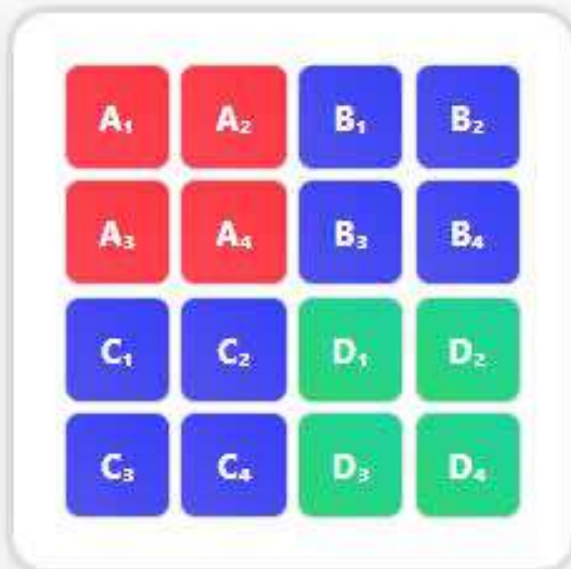
Паралельне збирання

Окрема пам'ять для кожного блоку

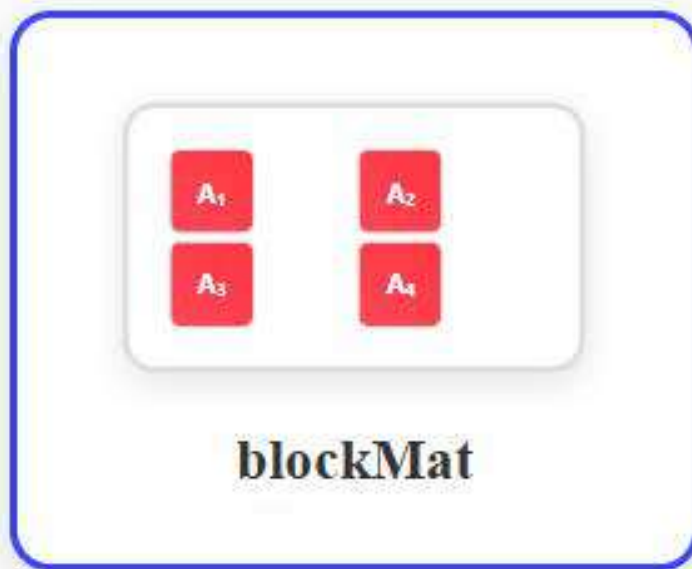
Обрахунок Доповнення Шура

Розріджена $A \rightarrow$ густа, щоб порахувати обернену через LU-факторизацію

Початкова матриця



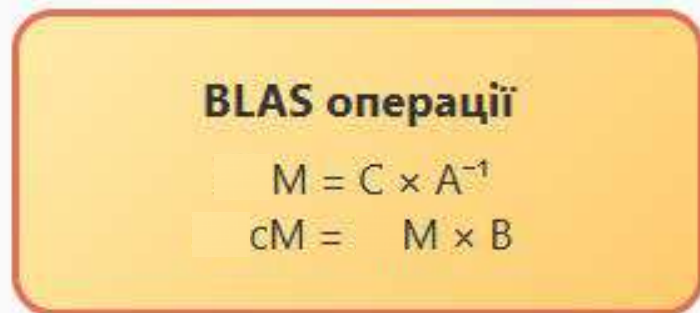
Збирання густої A



Обернення A



$C \cdot A^{-1} \cdot B$

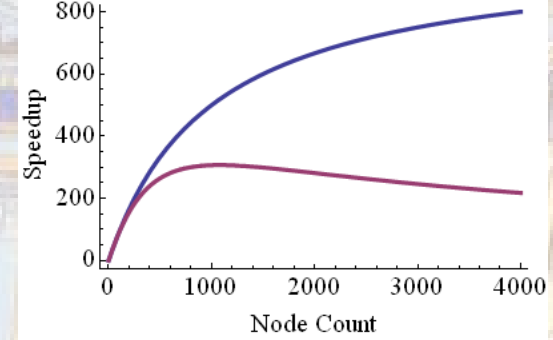


Доповнення Шура




Висновок

Закон Амдела



12/12

Результати досягнених в роботі прискорень наведені в таблиці:

Платформа	Лінійна алгебра	Загальне прискорення
 Archer2 <i>(Найбільший суперкомп'ютер Великої Британії)</i>	~12000% (120x)	~950% (9.5x)
Nvidia H100 <i>(\$30,000)</i>	~1000% (10x)	~230% (2.3x)
Nvidia GTX1050 <i>(бюджетна)</i>	~140% (1.4x)	~105% (1.05x)