

Bakalářská práce

Implementace maticových řešičů na GPU The implementation of matrix solvers on GPU

Martin Černý

školitel: Tomáš Oberhuber

České vysoké učení technické v Praze
Fakulta jaderná a fyzikálně inženýrská

31. srpna 2011

- 1 Popis algoritmů
- 2 Měření na CPU
- 3 Měření na GPU
- 4 Závěr

Metody řešení

- Sekvenční metoda
 - Thomasův algoritmus
- Paralelní metody
 - Cyklická metoda
 - Paralelní cyklická metoda

Sekvenční metoda

- Využívá Gaussovu eliminaci
 - Převod do horního trojúhelníkového tvaru
 - Převod na diagonální matici
- Nelze spustit paralelně
 - k výpočtu i -tého řádku je potřeba výsledek předchozího řádku

Elementární úprava

$$a_{i-1}x_{i-2} + b_{i-1}x_{i-1} + c_{i-1}x_i = h_{i-1}$$

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = h_i$$

$$a_{i+1}x_i + b_{i+1}x_{i+1} + c_{i+1}x_{i+2} = h_{i+1}$$

$$k_1 = \frac{a_i}{b_{i-1}} \quad k_2 = \frac{c_i}{b_{i+1}}$$

$$a_i^{(1)}x_{i-2} + b_i^{(1)}x_i + c_i^{(1)}x_{i+2} = h_i^{(1)}$$

$$a_i^{(1)} = -k_1 a_{i-1} \quad b_i^{(1)} = b_i - k_1 c_{i-1} - k_2 a_{i+1}$$

$$c_i^{(1)} = -k_2 c_{i+1} \quad h_i^{(1)} = h_i - k_1 h_{i-1} - k_2 h_{i+1}$$

Elementární úprava

$$a_{i-1}x_{i-2} + b_{i-1}x_{i-1} + c_{i-1}x_i = h_{i-1}$$

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = h_i$$

$$a_{i+1}x_i + b_{i+1}x_{i+1} + c_{i+1}x_{i+2} = h_{i+1}$$

$$k_1 = \frac{a_i}{b_{i-1}} \quad k_2 = \frac{c_i}{b_{i+1}}$$

$$a_i^{(1)}x_{i-2} + b_i^{(1)}x_i + c_i^{(1)}x_{i+2} = h_i^{(1)}$$

$$a_i^{(1)} = -k_1 a_{i-1} \quad b_i^{(1)} = b_i - k_1 c_{i-1} - k_2 a_{i+1}$$

$$c_i^{(1)} = -k_2 c_{i+1} \quad h_i^{(1)} = h_i - k_1 h_{i-1} - k_2 h_{i+1}$$

Elementární úprava

$$a_{i-1}x_{i-2} + b_{i-1}x_{i-1} + c_{i-1}x_i = h_{i-1}$$

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = h_i$$

$$a_{i+1}x_i + b_{i+1}x_{i+1} + c_{i+1}x_{i+2} = h_{i+1}$$

$$k_1 = \frac{a_i}{b_{i-1}} \quad k_2 = \frac{c_i}{b_{i+1}}$$

$$a_i^{(1)}x_{i-2} + b_i^{(1)}x_i + c_i^{(1)}x_{i+2} = h_i^{(1)}$$

$$a_i^{(1)} = -k_1 a_{i-1} \quad b_i^{(1)} = b_i - k_1 c_{i-1} - k_2 a_{i+1}$$

$$c_i^{(1)} = -k_2 c_{i+1} \quad h_i^{(1)} = h_i - k_1 h_{i-1} - k_2 h_{i+1}$$

Elementární úprava

$$a_{i-1}x_{i-2} + b_{i-1}x_{i-1} + c_{i-1}x_i = h_{i-1}$$

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = h_i$$

$$a_{i+1}x_i + b_{i+1}x_{i+1} + c_{i+1}x_{i+2} = h_{i+1}$$

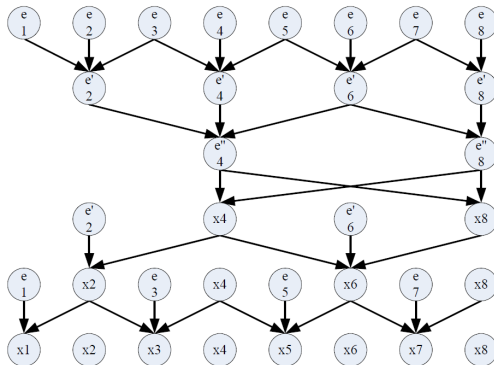
$$k_1 = \frac{a_i}{b_{i-1}} \quad k_2 = \frac{c_i}{b_{i+1}}$$

$$a_i^{(1)}x_{i-2} + b_i^{(1)}x_i + c_i^{(1)}x_{i+2} = h_i^{(1)}$$

$$a_i^{(1)} = -k_1 a_{i-1} \quad b_i^{(1)} = b_i - k_1 c_{i-1} - k_2 a_{i+1}$$

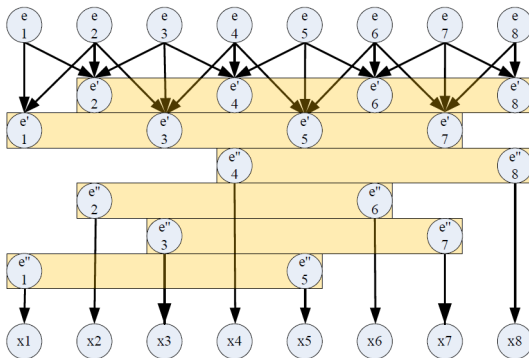
$$c_i^{(1)} = -k_2 c_{i+1} \quad h_i^{(1)} = h_i - k_1 h_{i-1} - k_2 h_{i+1}$$

Cyklická redukce



Obrázek: Zobrazení jednotlivých kroků při řešení soustavy s osmi neznámými algoritmem CR, e1 až e8 značí řádky matice, x1 až x8 jsou výsledky

Paralelní cyklická redukce



Obrázek: Zobrazení jednotlivých kroků při řešení soustavy s osmi neznámými algoritmem CR, e_1 až e_8 značí řádky matice, x_1 až x_8 jsou výsledky

Složitosti algoritmů

Tabulka: Složitosti používaných algoritmů

	Sekvenční	CR	PCR
kroků	$2n$	$2\log_2 n - 1$	$\log_2 n$
operací	$8n$	$17n$	$12n\log_2 n$

Paralelizace pomocí OpenMP

- používané metody
 - cyklická redukce
 - paralelní cyklická redukce
 - PCR střídavý způsob
 - PCR blokový způsob

Paralelizace pomocí OpenMP

- používané metody
 - cyklická redukce
 - paralelní cyklická redukce
 - PCR střídavý způsob
 - PCR blokový způsob

123...123...123...

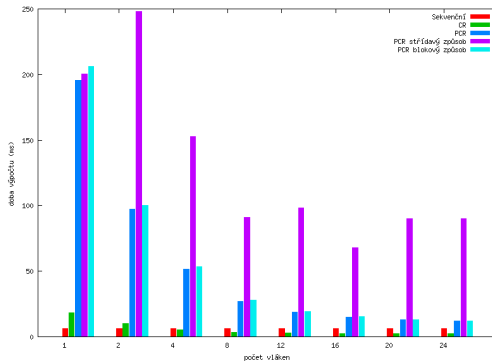
Paralelizace pomocí OpenMP

- používané metody
 - cyklická redukce
 - paralelní cyklická redukce
 - PCR střídavý způsob
 - PCR blokový způsob

123...123...123...

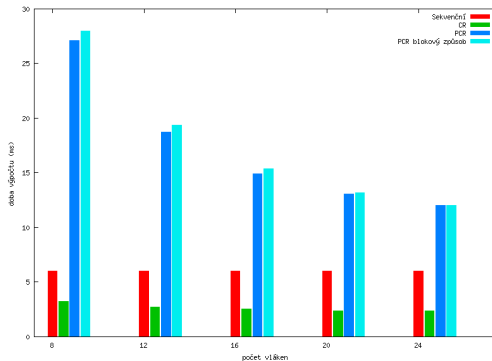
11...1 22...2 33...3 ...

Naměřené doby výpočtu



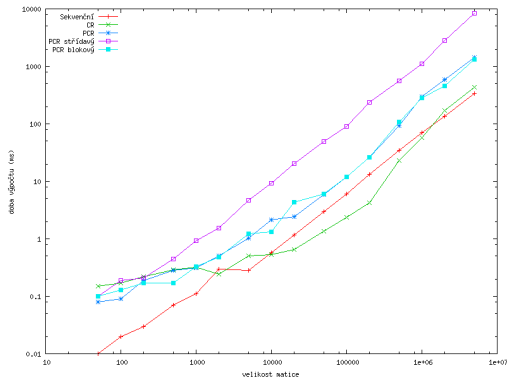
Obrázek: Doba výpočtu tridiagonální matice velikosti 100000

Naměřené doby výpočtu



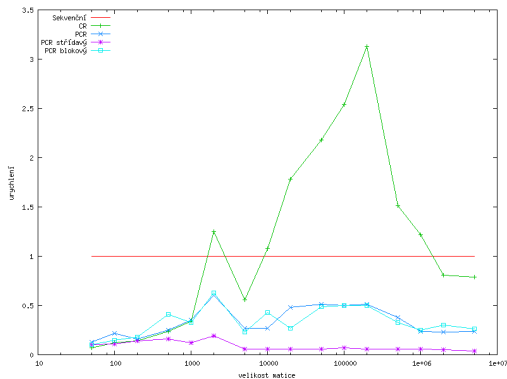
Obrázek: Doba výpočtu tridiagonální matice velikosti 100000

Naměřené doby výpočtu



Obrázek: Doby výpočtů jednotlivých metod, k dispozici je 24 jader

Naměřené doby výpočtu



Obrázek: Urychlení jednotlivých metod vzhledem k sekvenční, k dispozici je 24 jader

Výsledky

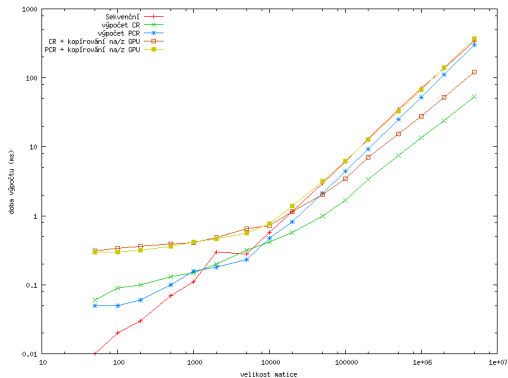
Tabulka: Doby výpočtů a urychlení jednotlivých metod na CPU vzhledem k sekvenční metodě

velikost matice	Sekvenční čas(ms)	CR		PCR		PCR střídavý		PCR blokový	
		čas(ms)	urychl	čas(ms)	urychl	čas(ms)	urychl	čas(ms)	urychl
50	0,01	0,15	0,07	0,08	0,13	0,1	0,1	0,1	0,1
100	0,02	0,17	0,12	0,09	0,22	0,19	0,11	0,13	0,15
200	0,03	0,22	0,14	0,19	0,16	0,21	0,14	0,17	0,18
500	0,07	0,29	0,24	0,28	0,25	0,44	0,16	0,17	0,41
1000	0,11	0,32	0,34	0,31	0,35	0,92	0,12	0,33	0,33
2000	0,3	0,24	1,25	0,5	0,6	1,55	0,19	0,48	0,63
5000	0,28	0,5	0,56	1,03	0,27	4,65	0,06	1,22	0,23
10000	0,57	0,53	1,08	2,14	0,27	9,32	0,06	1,33	0,43
20000	1,16	0,65	1,78	2,42	0,48	20,23	0,06	4,33	0,27
50000	2,94	1,35	2,18	5,81	0,51	49,32	0,06	6,05	0,49
100000	6,03	2,37	2,54	12	0,5	90,23	0,07	12,05	0,5
200000	13,1	4,19	3,13	25,85	0,51	234,12	0,06	26,35	0,5
500000	34,78	23,09	1,51	92,54	0,38	562,97	0,06	106,68	0,33
1000000	69,95	57,12	1,22	294,61	0,24	1099,87	0,06	282,59	0,25
2000000	136,84	168,62	0,81	584,11	0,23	2813,3	0,05	458,87	0,3
5000000	338,8	431,35	0,79	1432,52	0,24	8367,46	0,04	1325,1	0,26

Paralelizace na GPU

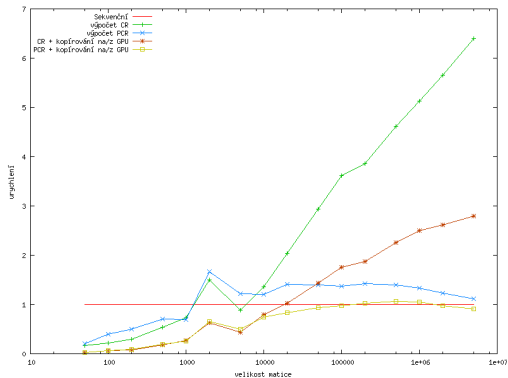
- používané metody
 - cyklická redukce
 - paralelní cyklická redukce

Naměřené doby výpočtu



Obrázek: Doby výpočtů jednotlivých metod

Naměřené doby výpočtu



Obrázek: Urychlení jednotlivých metod vzhledem k sekvenční

Výsledky

Tabulka: Doby výpočtů a urychlení jednotlivých metod na GPU vzhledem k sekvenční metodě

velikost matice	Sekvenční čas(ms)	pouze výpočet				výpočet + kopírování			
		CR		PCR		PCR		PCR	
		čas(ms)	urychl	čas(ms)	urychl	čas(ms)	urychl	čas(ms)	urychl
50	0,01	0,06	0,17	0,05	0,2	0,31	0,03	0,3	0,03
100	0,02	0,09	0,22	0,05	0,4	0,34	0,06	0,3	0,07
200	0,03	0,1	0,3	0,06	0,5	0,36	0,08	0,32	0,09
500	0,07	0,13	0,54	0,1	0,7	0,39	0,18	0,36	0,19
1000	0,11	0,15	0,73	0,16	0,69	0,41	0,27	0,42	0,26
2000	0,3	0,2	1,5	0,18	1,67	0,48	0,63	0,46	0,65
5000	0,28	0,32	0,88	0,23	1,22	0,65	0,43	0,56	0,5
10000	0,57	0,42	1,36	0,47	1,21	0,72	0,79	0,77	0,74
20000	1,16	0,57	2,04	0,82	1,41	1,14	1,02	1,39	0,83
50000	2,94	1	2,94	2,1	1,4	2,04	1,44	3,14	0,94
100000	6,03	1,67	3,61	4,39	1,37	3,43	1,76	6,15	0,98
200000	13,1	3,39	3,86	9,23	1,42	6,99	1,87	12,83	1,02
500000	34,78	7,55	4,61	24,8	1,4	15,49	2,25	32,74	1,06
1000000	69,95	13,64	5,13	52,45	1,33	27,97	2,5	66,78	1,05
2000000	136,84	24,16	5,66	111,25	1,23	52,2	2,62	139,29	0,98
5000000	338,8	52,92	6,4	301,99	1,12	121,62	2,79	370,69	0,91

Závěr

- Děkuji za pozornost
- Nějaké dotazy?

Chyba v měření na GPU

Tabulka: Chybné a správné měření času výpočtu na GPU

	CR		PCR	
	chybně	správně	chybně	správně
na GPU	1,21	1,24	1,25	1,23
výpočet	0,18	1,67	0,12	4,39
z GPU	2,1	0,62	4,9	0,61
celkem	3,49	3,53	6,27	6,23

Hardware

- CPU
2x AMD Opteron(tm) Processor 6172
- GPU
NVIDIA® GeForce® GTX 460

Ulítlé časy výpočtů na CPU

Tabulka: Cyklická redukce, jedno vlákno

velikost	jednotlivé časy výpočtů (ms)									
10	0,07	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
20	0,06	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02
50	0,08	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03
100	0,09	0,05	0,04	0,04	0,04	0,04	0,04	0,04	0,04	0,04
200	0,17	0,11	0,10	0,10	0,10	0,11	0,10	0,10	0,10	0,11
500	0,30	0,23	0,24	0,23	0,23	0,23	0,23	0,23	0,23	0,23
1000	0,51	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44
2000	0,54	0,51	0,51	0,50	0,50	0,51	0,51	0,51	0,51	0,51
5000	1,62	1,58	1,58	1,58	1,58	1,58	1,58	1,58	1,58	1,50
10000	4,04	1,64	1,64	1,64	1,64	1,64	1,64	1,64	1,64	1,64
20000	5,13	3,49	3,48	3,47	3,48	3,48	3,48	3,48	3,47	3,47
50000	13,84	9,61	9,70	8,95	8,99	9,06	9,03	9,02	9,01	8,97
100000	29,27	18,51	18,23	18,16	18,15	18,13	18,13	18,07	18,14	18,11
200000	45,97	46,50	46,41	46,48	46,47	46,43	46,45	46,45	46,46	46,42
500000	121,6	134,0	133,9	131,5	133,9	134,0	127,0	126,9	126,8	126,9
1000000	242,6	255,8	255,9	256,0	255,9	255,7	255,7	255,6	255,6	255,6
2000000	485,2	486,7	486,6	487,1	486,5	486,7	486,6	487,0	486,4	486,6
5000000	1292	1276	1254	1216	1234	1291	1299	1298	1291	1263
10000000	2420	2423	2419	2419	2420	2425	2428	2430	2438	2421

Ulítlé časy výpočtů na CPU

- AMD Opteron(tm) Processor 6172
- cache 12MB
- jeden řádek: $8 \times \text{double} = 64 \text{bytu}$
- do cache se vejde $\frac{12 \text{MB}}{64} = 196608$ řádků

Tabulka: Správnost při použití datového typu *float* ve výpisu 2.3, řádek 9

velikost matice	s r o v n á n í		
	SEK-CR	SEK-PCR	CR-PCR
N			
100	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
1000	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
10000	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
100000	102.0	102.0	SPRÁVNĚ
1000000	9646.0	9646.0	SPRÁVNĚ

Tabulka: Správnost při použití datového typu *double* ve výpisu 2.3, řádek 9

velikost matice	s r o v n á n í		
	SEK-CR	SEK-PCR	CR-PCR
N			
100	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
1000	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
10000	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
100000	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ
1000000	SPRÁVNĚ	SPRÁVNĚ	SPRÁVNĚ