# Московский Государственный Университет имени М. В. Ломоносова Факультет Вычислительной Математики и Кибернетики Кафедра Суперкомпьютеров и Квантовой Информатики

# Практикум CUDA

## Отчет №2

Basic Image Convolution on NVIDIA GPUs using CUDA(with performance improvements)

Работу выполнил **Кислов Евгений** 

#### Постановка задачи

Для выполнения второго задания необходимо реализовать следующие оптимизации разработанной на первом этапе программы:

Оптимизации для обработки больших и малых изображений:

- 1. Развертка массива, где хранится изображение из массива структур в структуру
- 2. массивов для улучшения шаблона доступа к глобальной памяти (Pixel \* -> 3 массива
- 3. unsigned char\* для хранения 3 компонент изображения);
- 4. Последовательный доступ к памяти от нитей варпа к массиву с изображением;
- 5. Использование разделяемой (shared) памяти для применения фильтра (по аналогии со stencil);
- 6. Использование 3х нитей для обработки r/g/b компонент;
- 7. Различные походы к передаче фильтра в матрицу (full unroll, константная память);
- 8. Развертка циклов, применяющих фильтров внутри каждой нити;
- 9. Подбор оптимальных значений размера CUDA блока;
- 10. Минимизация числа простаивающих нитей;

Дополнительные оптимизации для обработки набора из маленьких изображений:

- 1. Выделение памяти (cudaMalloc) под обрабатываемые изображения 1 раз (а не каждый раз для каждого изображения заново);
- 2. Обработка нескольких изображений за раз одним ядром или обработка нескольких
  - изображений в конкурентном режиме при помощи CUDA-потоков;
- 3. Одновременные копирования DtoH, HtoD и запуск ядер;
- 4. Параллельная работа с файлами обработка изображений на GPU для групп из N -

изображений: загружаем группу из N изображений с диска, пока их обрабатываем - грузим следующую. Сохранение на диск можно отключить (ifdef NEED\_TO\_SAVE\_\_).

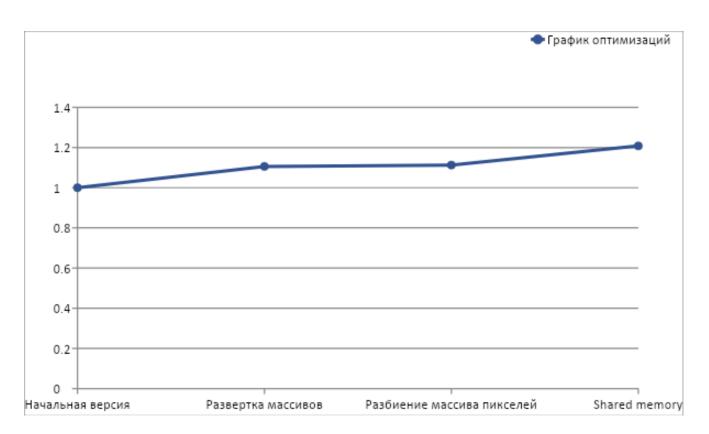
# Результат

Из списка общих оптимизаций были выполнены все пункты, из списка оптимизаций для маленький изображений было выполнены пункты 2-4.

Ниже будут приведены замеры времени и графики для наиболее важных оптимизаций:

Вид оптимизации	Время вычисления	Время вычисления	Ускорение(в
	на большой	на маленьких	среднем по двум
	картинке(ms)	картинках(ms)	случаям)
Начальная версия	0.602816	0.26464	-

Развертка массивов	0.54472	0.216864	1.1
Разбиение массива пикселей на три	0.541745	0.213487	1.1
Shared memory	0.498944	0.19292	1.22



#### Графики оптимизаций:

После оптимизаций маленьких картинок время пересылки уменьшилось с 4.83 мс до 3.64 мс, дав ускорение в 1.32 раза.

Данные профилировщика nvprof в начальной версии:

#### Для больших

==118650== Profiling result	+•												
Start Duration		Block Size	Reas*	SSMem*	DSMem*	Size	Throughout	SrcMemType	DstMemType	Device	Context	Stream	Name
484.03ms 3.2009ms			-				4.5250GB/s	Pageable		Tesla P100-SXM2	1		[CUDA memcpy HtoD]
487.49ms 896ns						36B	38.317MB/s	Pageable	Device	Tesla P100-SXM2			[CUDA memcpy HtoD]
487.73ms 550.79us	(90 57 1)	(32 32 1)	40	0B	0B					Tesla P100-SXM2			apply_kernel_device(unsigned char*
, unsigned char*, int, int,	, float∗, char) [22:	1]											
488.30ms 2.8804ms						14.832MB	5.0285GB/s	Device	Pageable	Tesla P100-SXM2	1	7	[CUDA memcpy DtoH]
Regs: Number of registers of SSMem: Static shared memory DSMem: Dynamic shared memory SrcMemType: The type of soo DstMemType: The type of dea	y allocated per CUD ry allocated per CUI urce memory accessed	A block. DA block. d by memory o	peration/co	ору	used in	ternally b	y the CUDA d	river and/or	tools and c	an be more than wh	at the com	piler sho	ws.

==116070== Profil:	ing resul	t:					
Type	Time(%)	Time	Calls	Avg	Min	Max	Name
GPU activities:	47.66%	3.1790ms	2	1.5895ms	896ns	3.1781ms	[CUDA memcpy HtoD]
	44.08%	2.9401ms	1	2.9401ms	2.9401ms	2.9401ms	[CUDA memcpy DtoH]
	8.26%	550.86us	1	550.86us	550.86us	550.86us	apply_kernel_device(unsigned char*, unsigned char*, int, in
t, float*, char)							
API calls:	94.06%	190.50ms	4	47.626ms	1.2870us	190.50ms	cudaEventCreate
	3.31%	6.7104ms	3	2.2368ms			cudaMemcpy
	1.03%	2.0892ms	2	1.0446ms		1.0500ms	cuDeviceTotalMem
	0.47%	949.17us	188	5.0480us	247ns	192.77us	cuDeviceGetAttribute
		561.87us	4	140.47us	4.1210us		cudaEventSynchronize
		552.57us	3		156.97us	223.64us	cudaMalloc
		517.44us	3	172.48us		248.28us	cudaFree
		474.06us	1	474.06us			cudaGetDeviceProperties
	0.04%	79.603us	2	39.801us		41.162us	cuDeviceGetName
		44.633us	1	44.633us	44.633us	44.633us	cudaLaunch
		32.840us	4		5.0920us	15.635us	cudaEventRecord
		9.8810us	2	4.9400us	4.8670us	5.0140us	
		2.6830us	3	894ns	385ns	1.8980us	cuDeviceGetCount
		1.9580us	6	326ns	258ns	444ns	cudaSetupArgument
		1.5800us	4	395ns	246ns	652ns	cuDeviceGet
	0.00%	1.3930us	1	1.3930us	1.3930us	1.3930us	cudaConfigureCall

## Для маленьких

==116853== Profiling result:														
Start Duration	Grid Size	Block Size	Regs*	SSMem* I	DSMem*				DstMemType	Device	Context	Stream		
270.48ms 11.040us						263.67KB	22.777GB/s	Pageable	Device	Tesla P100-SXM2		7	[CUDA memcp:	y HtoD]
270.50ms 544ns						36B	63.111MB/s	Pageable	Device	Tesla P100-SXM2			[CUDA memcp	y HtoD]
270.57ms 11.552us	(10 10 1)	(32 32 1)	40	ØB	0B					Tesla P100-SXM2	1		apply_kerne	l_device(unsigned char*
, unsigned char*, int, int,	float*, char)	[221]												
270.60ms 12.001us						263.67KB	20.953GB/s	Device	Pageable	Tesla P100-SXM2			[CUDA memcp	v DtoH1
309.67ms 10.401us							24.176GB/s	Pageable		Tesla P100-SXM2	1		[CUDA memcp	
309.69ms 512ns							67.055MB/s	Pageable		Tesla P100-SXM2			[CUDA memcp	
309.73ms 11.105us	(10 10 1)	(32 32 1)	40	ØB.	θВ	-	-	- uguubiu		Tesla P100-SXM2	ī			l_device(unsigned char*
, unsigned char*, int, int,				00	00					16318 F100-3AM2			appry_kerile.	1_device(disigned char+
309.76ms 12.001us	iluat*, char,	[202]				0/0 /7/0	20.953GB/s	Device	0	Tesla P100-SXM2	1		[CUDA memcp	. 04-113
348.65ms 10.369us							24.251GB/s	Pageable		Tesla P100-SXM2	1		[CUDA memcp	
348.67ms 480ns						368	71.526MB/s	Pageable		Tesla P100-SXM2			[CUDA memcp	
348.72ms 10.816us	(10 10 1)	(32 32 1)	40	0B	0B					Tesla P100-SXM2			apply_kerne:	l_device(unsigned char*
, unsigned char*, int, int,		[283]												
348.74ms 11.969us							21.009GB/s	Device		Tesla P100-SXM2			[CUDA memcp:	
388.79ms 10.176us						263.67KB	24.711GB/s	Pageable	Device	Tesla P100-SXM2		7	[CUDA memcp:	y HtoD]
388.81ms 512ns						36B	67.055MB/s	Pageable	Device	Tesla P100-SXM2			[CUDA memcp:	y HtoD]
388.85ms 10.753us	(10 10 1)	(32 32 1)	40	0B	0B					Tesla P100-SXM2			apply_kerne	l_device(unsigned char*
, unsigned char*, int, int,	float*, char)	[314]												
388.88ms 12.353us						263.67KB	20.356GB/s	Device	Pageable	Tesla P100-SXM2	1		[CUDA memcp	v DtoH1
426.11ms 9.9840us							25.186GB/s	Pageable		Tesla P100-SXM2	1		[CUDA memcp	
426.13ms 512ns							67.055MB/s	Pageable		Tesla P100-SXM2	1		[CUDA memcp	
426.17ms 10.785us	(10 10 1)	(32 32 1)	40	<b>0</b> B	θВ	-	-	- ugcubic		Tesla P100-SXM2	ī			l_device(unsigned char*
, unsigned char*, int, int,				00	0.0					16318 F100-3AM2			apply_kerne.	1_device(unsigned char-
426.20ms 12.033us	Troute, char,	[540]				242 4740	20.897GB/s	Device	Dagaahla	Tesla P100-SXM2	1		[CUDA memcp	D+oU1
462.69ms 10.048us							25.026GB/s	Pageable		Tesla P100-SXM2	1		[CUDA memcp	
462.71ms 512ns		(00 00 4)		-		308	67.055MB/s	Pageable		Tesla P100-SXM2			[CUDA memcp	
462.75ms 10.817us	(10 10 1)	(32 32 1)	40	0B	<b>0</b> B					Tesla P100-SXM2			appiy_kerne.	l_device(unsigned char*
, unsigned char*, int, int,	float*, char)	[376]												
462.78ms 12.032us							20.899GB/s	Device		Tesla P100-SXM2			[CUDA memcp	
512.94ms 10.049us							25.023GB/s	Pageable		Tesla P100-SXM2			[CUDA memcp:	
512.96ms 512ns							67.055MB/s	Pageable		Tesla P100-SXM2			[CUDA memcp:	
513.00ms 10.784us	(10 10 1)	(32 32 1)	40	0B	0B					Tesla P100-SXM2			apply_kerne	l_device(unsigned char*
, unsigned char*, int, int,	float*, char)	[407]												
513.03ms 12.065us						263.67KB	20.842GB/s	Device	Pageable	Tesla P100-SXM2	1		[CUDA memcp:	y DtoH]
Regs: Number of registers us SSMem: Static shared memory DSMem: Dynamic shared memory SrcMemType: The type of sour DstMemType: The type of desi	allocated per allocated per ce memory acc	CUDA block. r CUDA block. essed by memory	operation/co	ору	used ir	nternally b	y the CUDA o	river and/o	r tools and c	an be more than w	hat the com	piler sho	ows.	
				.,		· ·	,						,	
==117457== Profili	na result	•												
Type	Time(%)	Time	Calls	<i>F</i>	٩vg	Mi	.n	Max Nar	ne					
GPU activities:	35 //3%	84.197us	_7			12 000	ıs 12.00	Stue LCI	JDA memcr	ov DtoH1				
oro activities:														
	32.38%	76.932us	7	10.996	∂us	10.785ι	ıs 11.63	Lous app	ply kerne	el device(un	signed	char*,	unsigned	d char*, int, in
+ flooty obox)														
t, float*, char)														
	22 109/	74 /07	1 /	E //20	3110	E12-	0 11 0/	10.10	IDA momor	U+oD1				

II/45/ PIUIII	TIIR TESUT						
Type	Time(%)	Time	Calls	Avg	Min	Max	Name
GPU activities:	35.43%	84.197us	7	12.028us	12.000us	12.065us	[CUDA memcpy DtoH]
	32.38%	76.932us	7	10.990us	10.785us	11.616us	apply_kernel_device(unsigned char*, unsigned char*, int, in
t, float*, char)							
	32.19%	76.487us	14	5.4630us	512ns	11.809us	[CUDA memcpy HtoD]
API calls:	95.15%	139.21ms	28	4.9718ms	1.2970us	139.15ms	cudaEventCreate
	1.43%	2.0897ms	2	1.0449ms	1.0387ms	1.0511ms	cuDeviceTotalMem
	0.87%	1.2776ms	21	60.839us	7.4290us	172.59us	cudaMalloc
	0.73%	1.0720ms	21	51.046us	8.5000us	147.62us	cudaFree
	0.69%	1.0145ms	188	5.3960us	241ns	215.63us	cuDeviceGetAttribute
	0.41%	601.36us	21	28.636us	9.1310us	77.857us	cudaMemcpy
	0.32%	469.19us	1	469.19us	469.19us	469.19us	cudaGetDeviceProperties
	0.11%	161.31us	28	5.7610us	3.5550us	11.061us	cudaEventSynchronize
	0.09%	132.79us	7	18.969us	14.802us	38.088us	cudaLaunch
	0.08%	114.29us	28	4.0810us	3.3830us	10.508us	cudaEventRecord
	0.06%	85.042us	2	42.521us	38.385us	46.657us	cuDeviceGetName
	0.04%	64.138us	14	4.5810us	4.2270us	5.5800us	cudaEventElapsedTime
	0.01%	12.617us	42	300ns	242ns	1.0160us	cudaSetupArgument
	0.00%	3.4250us	7	489ns	344ns	920ns	cudaConfigureCall
	0.00%	2.6510us	3	883ns	367ns	1.8700us	cuDeviceGetCount
	0.00%	_1.5290us	4	382ns	257ns	571ns	cuDeviceGet

# Данные профилировщика nvprof в оптимизорованной версии:

## Для больших

```
==186956em Profiling application: /e.out emboss —b input_img_big/big_1.png output_img_big/big_1.png
==186956em Profiling application: /e.out emboss —b input_img_big/big_1.png output_img_big/big_1.png

Start Duration

489.08ms 928ns Grid Size Block Size Regs* SSMes* DSMes* Size Throughput SrcMemType DstWemType Device Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

489.08ms 97.12us — — — — — 4.9438MB 4.928ms Profiling Regular Device Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

499.48ms 965.59us — — — — 4.9438MB 5.988808/8 Pageable Device Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

409.48ms 965.59us — — — 4.9438MB 6.988808/8 Pageable Device Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

409.48ms 965.59us — — 4.9438MB 6.988808/8 Pageable Device Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

409.48ms 965.59us — — 4.9438MB 6.988808/8 Pageable Device Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

409.48ms 965.59us — — 4.9438MB 6.9837808/8 Pageable Device Pageable Tesla plae-SAMZ 1 15 [CUDA memcpy HtdD]

409.28ms 1.27271ms — — 4.9438MB 6.9337808/8 Device Pageable Tesla plae-SAMZ 1 15 [CUDA memcpy DtdH]

409.28ms 1.27271ms — — 4.9438MB 6.9337808/8 Device Pageable Tesla plae-SAMZ 1 15 [CUDA memcpy DtdH]

409.28ms 1.818.66us — — — 4.9438MB 6.9337808/8 Device Pageable Tesla plae-SAMZ 1 15 [CUDA memcpy DtdH]

88gs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler shows.

88Mem: Static shared memory allocated per CUDA block.

89Mem: Static shared memory allocated per CUDA block.

89Mem: Pageable Tesla plae-SAMZ 1 15 [CUDA memcpy DtdH]

89Extended per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler shows.
```

```
### Profiling result:
Type Time(%) Time (%) Time
```

#### Для маленьких

	= Profiling																	
Start 271.90ms	Duration 929ns	Gr	id Size	Block Size	Regs*	SSMem* DS	Mem∗	Size 36B	Throughput 36.956MB/s	Pageable		Device Tesla P100-SXM2	Context 1	Stream 15	Name	memcpy	U+oD1	
	4.2880us						_ a		19.547GB/s	Pageable		Tesla P100-SXM2	i	15		memcpy		
	4.0330us								20.783GB/s	Pageable		Tesla P100-SXM2	ī	15		memcpy		
	4.1610us								20.144GB/s	Pageable		Tesla P100-SXM2	ī	15		memcpy		
	11.105us	(1)	0 10 1)	(32 32 1)	36 3	.3867KB	0B					Tesla P100-SXM2		15			device_3x3(un	signed c
		unsigned c	har∗, unsigne	ed char∗, uns:	igned char∗	, unsigned ch			float*) [231]									-
	3.7760us								22.198GB/s	Device		Tesla P100-SXM2		15		memcpy		
	3.7120us								22.581GB/s	Device		Tesla P100-SXM2	1	15		memcpy		
272.11ms 311.73ms	3.6160us 704ns						- 8 -	87.891KB 36B	23.180GB/s 48.767MB/s	Device Pageable		Tesla P100-SXM2 Tesla P100-SXM2	1	15 16		memcpy memcpy		
	3.8730us								21.642GB/s	Pageable		Tesla P100-SXM2	1	16		memcpy		
	3.6800us								22.777GB/s	Pageable		Tesla P100-SXM2	i	16		memcpy		
	4.0000us								20.955GB/s	Pageable		Tesla P100-SXM2	ī	16		memcpy		
	10.593us	(1)	9 10 1)	(32 32 1)	36 3	.3867KB	0В					Tesla P100-SXM2	ī	16			device_3x3(un	signed c
har*, uns	igned char*,	unsigned c	har∗, unsigne	ed char*, uns:	igned char*	, unsigned ch	ar∗, in	nt, int,	float*) [280]									
	3.7120us								22.581GB/s	Device		Tesla P100-SXM2		16		memcpy		
	3.6800us								22.777GB/s	Device		Tesla P100-SXM2		16		memcpy		
	3.6170us								23.174GB/s	Device		Tesla P100-SXM2		16		memcpy		
350.96ms	704ns									Pageable		Tesla P100-SXM2	1	17		memcpy		
	3.8080us								22.011GB/s	Pageable		Tesla P100-SXM2	1	17		memcpy		
	3.8720us 3.8080us								21.647GB/s 22.011GB/s	Pageable Pageable	Device	Tesla P100-SXM2 Tesla P100-SXM2	1	17 17		memcpy memcpy		
	10.369us	(1)	9 10 1)	(32 32 1)		.3867KB	9B	27.03TVD	22.01105/8	Pageable _		Tesla P100-SXM2	1	17			device_3x3(un	eigned c
								nt. int.	float*) [329]			16314 F100-3AMZ			abbia-	Kermer_	device_5x3(u)	isigned c
	3.7120us				-	-			22.581GB/s	Device	Pageable	Tesla P100-SXM2	1	17	[CUDA	memcpy	DtoH]	
351.09ms	3.6800us						- 8	87.891KB	22.777GB/s	Device	Pageable	Tesla P100-SXM2		17	[CUDA	memcpy	DtoH]	
351.11ms	3.6810us						- 8		22.771GB/s	Device	Pageable	Tesla P100-SXM2			[CUDA	memcpy	DtoH]	
391.51ms	704ns								48.767MB/s	Pageable		Tesla P100-SXM2		18		memcpy		
391.52ms	3.7760us																	
									22.198GB/s	Pageable	Device	Tesla P100-SXM2	1	18	[CUDA	memcpy	HTOD]	
	48== Profi 48== Profi	ling resul		/a.out embo			all ou	utput_i	mg_small	Pageable	Device	Tesla P100-SXM2	1	18	[CUDA	memcpy	HEOD]	
==10244	48== Profi 48== Profi Type	ling resul Time(%)	lt: Time	/a.out embo Calls	Avg	Min	all ou	utput_i Max	mg_small	-	Device	Tesla P100-SXM2	1	18	[ CUDA	memcpy	HEODJ	
==10244	48== Profi 48== Profi	ling resul Time(%) 37.13%	t: Time 89.028us	/a.out embo Calls 28	Avg 3.1790us	Min 704ns	all ou	utput_i Max 480us	mg_small Name [CUDA memcp	y HtoD]	Device	Tesla P100-SXM2	1	18	[CUDA	memcpy	HEODJ	
==10244	48== Profi 48== Profi Type	ling resul Time(%) 37.13% 32.17%	Time 89.028us 77.121us	/a.out embo Calls 28 21	Avg 3.1790us 3.6720us	Min 704ns 3.6160us	5.24 3.74	utput_i Max 480us 440us	mg_small Name [CUDA memcp [CUDA memcp	y HtoD] oy DtoH]								
==10244 GPU ac	48== Profi 48== Profi Type ctivities:	ling resul Time(%) 37.13% 32.17% 30.70%	Time 89.028us 77.121us 73.605us	/a.out embo Calls 28 21 7	Avg 3.1790us 3.6720us 10.515us	Min 704ns 3.6160us	5.24 3.74	utput_i Max 480us 440us	mg_small Name [CUDA memcp [CUDA memcp	y HtoD] oy DtoH]		Tesla P100-SXM2						char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c	t: Time 89.028us 77.121us 73.605us char*, int,	/a.out embo Calls 28 21 7 , int, floa	Avg 3.1790us 3.6720us 10.515us t*)	Min 704ns 3.6160us 10.369us	5.24 3.74 10.9	utput_i Max 480us 440us 944us	mg_small Name [CUDA memcp [CUDA memcp apply_kerne	oy HtoD] oy DtoH] ol_device_								char*,
GPU ac	48== Profi 48== Profi Type ctivities:	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c	Time 89.028us 77.121us 73.605us char*, int, 139.14ms	Calls 28 21 7 , int, floa	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms	Min 704ns 3.6160us 10.369us	5.24 3.74 10.9	utput_i Max 480us 440us 944us	mg_small Name [CUDA memcp [CUDA memcp apply_kerne	oy HtoD] oy DtoH] el_device_ create								char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned of 94.39% 1.42%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms	Calls 28 21 7 , int, floa 7 2	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms	Min 704ns 3.6160us 10.369us 9.2850us	5.24 3.74 10.9	utput_i Max 480us 440us 944us .07ms 471ms	mg_small Name [CUDA memcp [CUDA memcp apply_kerne cudaStreamC cuDeviceTot	oy HtoD] oy DtoH] el_device_ create								char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.07%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms 1.5806ms	Calls 28 21 7 , int, floa 2 49	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us	Min 704ns 3.6160us 10.369us 9.2850us 1.0390ms 6.6210us	5.24 3.74 10.9 139.	utput_i Max 480us 440us 944us .07ms 471ms .44us	mg_small  Name [CUDA memcp [CUDA memcp apply_kerne cudaStreamC cuDeviceTot cudaMalloc	oy HtoD] oy DtoH] el_device_ create								char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned of 94.39% 1.42%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms 1.5806ms	Calls 28 21 7 , int, floa 7 2	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us	Min 704ns 3.6160us 10.369us 9.2850us	5.24 3.74 10.9 139.	utput_i Max 480us 440us 944us .07ms 471ms .44us	mg_small Name [CUDA memcp [CUDA memcp apply_kerne cudaStreamC cuDeviceTot	oy HtoD] oy DtoH] el_device_ create								char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.07%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms	Calls 28 21 7 , int, floa 2 49	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us	Min 704ns 3.6160us 10.369us 9.2850us 1.0390ms 6.6210us 7.8990us	5.24 3.74 10.9 139. 1.04 214.	utput_i Max 480us 440us 944us .07ms 471ms .44us .96us	mg_small Name [CUDA memcp [CUDA memcp apply_kerne cudaStreamC cuDeviceTot cudaMalloc	oy HtoD] oy DtoH] ol_device_ create calMem	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned 94.39% 1.42% 1.07% 0.94%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us	Calls 28 21 7 , int, floa 49 49	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.259us 5.1600us	Min 704ns 3.6160us 10.369us 9.2850us 1.0390ms 6.6210us 7.8990us	5.24 3.74 10.5 139. 1.04 214. 148.	Max 480us 440us 944us .07ms 471ms .44us .96us	mg_small  Name [CUDA memc; [CUDA memc; apply_kerne cudaStreamC cuDeviceTot cudaMalloc cudaFree cuDeviceGet	oy HtoD] by DtoH] cl_device_ create calMem	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.07% 0.94% 0.66% 0.65%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms 1.3847ms 970.18us 965.34us	/a.out embo  Calls 28 21 7 , int, floa 7 2 2 49 49 188 49	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.259us 5.1600us 19.700us	Min 704ns 3.6160us 10.369us 9.2850us 1.0390ms 6.6210us 7.8990us 246ns 10.254us	5.24 3.74 10.5 139. 1.04 214. 148. 201.	Max 480us 440us 944us 944us .07ms .471ms .44us .96us .95us 531us	mg_small  Name [CUDA memc; [CUDA memc; cudaStream( cuDeviceTot cudaMalloc cudaFree cuDeviceGet cudaMemcpyA	oy HtoD] by DtoH] cl_device_ create calMem cAttribute	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned of 94.39% 1.42% 1.07% 0.66% 0.66% 0.31%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.08661ms 1.5806ms 1.3847ms 970.18us 460.25us	/a.out embo  Calls 28 21 7 , int, floa 7 2 49 49 188 49	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.259us 5.1600us 19.700us 460.25us	Min 704ns 3.6160us 10.369us 9.2850us 1.0390ms 6.6210us 7.8990us 246ns 10.254us 460.25us	5.24 3.74 10.9 139. 1.04 214. 148. 201. 49.5	Max 480us 440us 944us .07ms 471ms .44us .96us .95us 531us .25us	mg_small  Name [CUDA memc; [CUDA memc; apply_kerne cudaStream( cudeStream( cudaMalloc cudaFree cuDeviceGet cudaMemcpy/ cudaGetDevi cudaGetDevi	oy HtoD] by DtoH] cl_device_ create calMem cAttribute	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.07% 0.94% 0.65% 0.65% 0.31% 0.12%	Time 89.028us 77.121us 73.605us char*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us 965.34us 460.25us 172.57us	Calls 28 21 7 , int, floa 49 49 188 49	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.257us 5.1600us 19.700us 460.25us 24.652us	Min 794ns 3.6160us 10.369us 9.2850us 1.0390ms 6.6210us 7.8990us 246ns 10.254us 460.25us 18.125us	5.24 3.74 10.9 139. 1.04 214. 148. 201. 49.5	Max 480us 440us 944us .07ms 471ms .44us .96us .95us 531us .25us	mg_small  Name [CUDA memc; [CUDA memc; apply_kerne cudaStream cuDeviceTot cudaMalloc cudaFree cudeFree cudewececudeGet cudaMemcpy/ cudaCatunch	oy HtoD] by DtoH] cl_device_ create calMem cAttribute sync ceePropert	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.07% 0.94% 0.66% 0.65% 0.31% 0.12%	Time 89.028us 77.121us 73.605us rhar*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us 965.34us 460.25us 172.57us 171.38us	/a.out embo  Calls 28 21 7 , int, floa 49 49 188 49 1 7 7 28	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.257us 5.1600us 19.700us 460.25us 24.652us 6.1200us	Min 704ns 3.6160us 10.369us 1.0390ms 6.6210us 2.2940ns 10.254us 460.25us 18.125us 3.6530us	5.24 3.74 10.5 139. 1.04 214. 148. 201. 49.5 460.	Max 480us 440us 944us .07ms 471ms .44us .96us .95us .25us 448us	Mg_small Name [CUDA memcp [CUDA memcp apply_kerne cudaStream cudewiceTot cudaMalloc cudaFree cuDeviceGet cudaMemcpy/ cudaGetDevi cudaEventS)	y HtoD] y DtoH] cl_device_ create calMem cAttribute sync cePropert	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.07% 0.66% 0.66% 0.65% 0.31% 0.12% 0.12%	89.028us 77.121us 73.605us 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us 965.34us 460.25us 172.57us 171.38us 141.91us	/a.out embo  Calls 28 21 7 , int, floa 7 2 49 49 188 49 1 7 28 28	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 5.1600us 19.700us 460.25us 24.652us 6.1200us 5.0680us	Min 784ns 3.6160us 18.369us 19.2850us 6.6210us 7.8990us 246ns 19.254us 460.25us 18.125us 3.6530us 3.9080us 3.9080us	5.24 3.74 10.5 139. 1.04 214. 148. 201. 49.5 460.4 11.6	Max 480us 440us 944us .07ms 471ms .44us .96us .95us 531us .25us 448us 004us 922us	mg_small  Name [CUDA memc; [CUDA memc; apply_kerne cudaStream( cuDeviceTot cudaMalloc cudaFree cuDeviceGet cudaMemcpy/ cudaGetDevi cudaGetDevi cudaLaunch cudaEventS; cudaEventRe	yy HtoD] yy DtoH] il_device create calMem  Attributo sync ceProperi	_3x3(unsig							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% 94.39% 1.42% 1.07% 0.66% 0.65% 0.31% 0.12% 0.12% 0.10%	89.028us 77.121us 73.605us 73.605us 73.14ms 2.0861ms 1.5806ms 1.5806ms 1.3847ms 970.18us 965.34us 460.25us 172.57us 171.38us 141.91us 91.496us	/a.out embo  Calls 28 21 7 , int, floa 49 49 188 49 1 7 28 28	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 5.1600us 19.700us 460.25us 6.1200us 5.0680us 45.748us	Min 794ns 3.6160us 10.369us 10.369us 2.850us 2.46ns 2.46ns 18.125us 18.125us 3.6530us 44.773us 44.773us	5.24 3.74 10.5 139. 1.04 214. 148. 201. 49.5 460. 40.4	Max 480us 440us 944us .07ms 471ms .44us .96us .531us .25us 448us 0004us 922us 723us	mg_small  Name [CUDA memc; [CUDA memc; apply_kerne cudaStream cudaMalloc cudaFree cubeviceGet cudaMemcpy/ cudaGetDevi cudaLunch cudaEventS; cudeEventS; cudeVecGet	yy HtoD] yy DtoH] cl_device. create calMem  Attribute csync ccePropert ccord Name	_3x3(unsig							char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned of 94.39% 1.42% 0.65% 0.65% 0.31% 0.12% 0.12% 0.10% 0.06%	89.028us 77.121us 73.605us rhar*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us 460.25us 172.57us 171.38us 141.91us 91.496us 68.107us	/a.out embo  Calls 28 21 7 , int, floa 7 2 49 49 188 49 1 7 28 28 2	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.259us 5.1600us 19.700us 460.25us 24.652us 6.1200us 5.0680us 45.748us 9.7290us	Min 784ns 3.6160us 10.369us 10.369us 6.6210us 7.8990us 246ns 10.254us 18.125us 3.6530us 3.9080us 44.773us 9.0130us 9.0130us	139. 1.04 214. 148. 201. 49.5 460. 40.4 11.6 10.5	utput_i  Max 480us 440us 944us .07ms .471ms .44us .95us .95us .25us .448us 004us 9723us 78703	mg_small Name [CUDA memcp [CUDA memcp apply_kerne cudextream( cuDeviceTot cudaMalloc cudaFree cuDeviceGet cudaMemcpyd cudaGetDevi cudaGetDevi cudaLaunch cudaEventRe cuDeviceGet cudaEventRe cuDeviceGet	by HtoD] by DtoH] cl_device. create alMem cattribute sync ccePropert cord cord Name lestroy	,3x3(unsign							char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% 94.39% 1.42% 1.07% 0.66% 0.65% 0.31% 0.12% 0.12% 0.10%	89.028us 77.121us 73.605us rhar*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us 460.25us 172.57us 171.38us 141.91us 91.496us 68.107us	/a.out embo  Calls 28 21 7 , int, floa 49 49 188 49 1 7 28 28	Avg 3.1790us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.259us 5.1600us 19.700us 460.25us 24.652us 6.1200us 5.0680us 45.748us 9.7290us	Min 794ns 3.6160us 10.369us 10.369us 2.850us 2.46ns 2.46ns 18.125us 18.125us 3.6530us 44.773us 44.773us	139. 1.04 214. 148. 201. 49.5 460. 40.4 11.6 10.5	utput_i  Max 480us 440us 944us .07ms .471ms .44us .95us .95us .25us .448us 004us 9723us 78703	mg_small  Name [CUDA memc; [CUDA memc; apply_kerne cudaStream cudaMalloc cudaFree cubeviceGet cudaMemcpy/ cudaGetDevi cudaLunch cudaEventS; cudeEventS; cudeVecGet	by HtoD] by DtoH] cl_device. create alMem cattribute sync ccePropert cord cord Name lestroy	,3x3(unsign							char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.67% 0.66% 0.31% 0.12% 0.12% 0.12% 0.10%	89.028us 77.121us 73.605us rhar*, int, 139.14ms 2.0861ms 1.5806ms 1.3847ms 970.18us 460.25us 172.57us 171.38us 141.91us 91.496us 68.107us	/a.out embo  Calls 28 21 7 , int, floa 49 49 188 49 1 7 7 28 28 28 7	Avg 3.1790us 3.6720us 10.515us t*) 119.877ms 1.0431ms 32.257us 28.259us 460.25us 460.25us 460.25us 5.1680us 5.1680us 45.748us 9.7290us 4.8220us	Min 784ns 3.6160us 10.369us 10.369us 6.6210us 7.8990us 246ns 10.254us 18.125us 3.6530us 3.9080us 44.773us 9.0130us 9.0130us	139. 139. 1.04 214. 148. 201. 49.5 460. 40.4 11.6 10.5	Max 480us 440us 944us 944us .07ms 471ms .44us .95us .95us .25us 448us 904us 922us 723us 787us	mg_small Name [CUDA memcp [CUDA memcp apply_kerne cudextream( cuDeviceTot cudaMalloc cudaFree cuDeviceGet cudaMemcpyd cudaGetDevi cudaGetDevi cudaLaunch cudaEventRe cuDeviceGet cudaEventRe cuDeviceGet	by HtoD] by DtoH] cl_device_ create calMem  Attribute c.cePropert crchronize cord Name cestroy casedTime	,3x3(unsign							char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time(%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 1.67% 0.66% 0.31% 0.12% 0.12% 0.12% 0.10%	Time 89.028us 77.121us 73.605us 73.605us 139.14ms 2.0861ms 1.5806/ms 2.08625us 1.5806/ms 460.25us 172.57us 171.38us 141.91us 91.496us 68.107us 67.520us 49.272us	/a.out embo  Calls 28 21 7 , int, floa 49 49 188 49 1 7 7 28 28 28 7	Avg 3.1779us 3.6720us 10.515us t*) 19.877ms 1.0431ms 32.257us 28.257us 19.700us 460.25us 24.652us 6.120us 5.0680us 4.5.748us 9.7290us 4.8220us 1.7590us	Min 704ns 3.6160us 10.369us 1.0399ms 6.6210us 7.8990us 246ns 12.25us 18.125us 3.9080us 44.773us 9.0130us 4.5020us	139. 1.04 214. 148. 201. 49.5 460. 40.4 11.6 10.5 46.7 11.7	Max 480us 440us 944us 944us 944us .07ms .441s .95us 531us .25us 448us 902us 723us 787us 440us	mg_small  Name [CUDA memcp[CUDA memcp] apply_kerne cudaStream cuDeviceTot cudaMalloc cudaFree cuDeviceGet cudaMemcpy/ cudaGetDevi cudaLaunch cudaEventS) cudaEventS cudeStream cudeViceGet cudaStream cudeStream cudaStream	yy HtoD] yy DtoH] tl_device. treate talmem  Attribute tsync tcePropert trectord Name lestroy tested	.3x3(unsigi ) :ties							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time (%) 37.13% 32.17% 38.70% unsigned c 94.39% 1.42% 1.67% 0.66% 0.65% 0.12% 0.12% 0.12% 0.10% 0.06% 0.06%	Time 89.028us 77.121us 73.605us 73.605us rhar*, int, 139.14ms 1.5806ms 1.5804ms 970.18us 970.18us 970.18us 171.38us 172.57us 171.38us 174.91us 67.52eus 68.107us 67.52eus 44.251us	/a.out embo  Calls 28 21 7, int, floa 49 49 188 49 1 7 7 28 28 28 29 14 28 7	Avg 3.1790us 3.6720us 10.515us t**) 19.877ms 1.0431ms 32.257us 28.259us 5.1660us 19.700us 460.25us 4.652us 6.1260us 5.0680us 4.5.748us 9.7290us 4.8220us 1.7590us 5.8930us	Min 784ns 3.6160us 10.369us 9.2850us 2.60210us 2.46ns 10.254us 460.25us 18.125us 3.6530us 3.9880us 4.73us 9.0130us 4.5020us 1.2770us 5.6690us 5.6690us 5.6690us 6.659us 6.659u	139. 1.04 214. 149. 201. 49. 460. 40.4 11.6 11.7 5.26 4.34 6.56	Max 480us 440us 444us 944us 975us 471ms 471ms 471ms 950us 951us 950us 448us 9723us 7723us 7723us 787us 960us 960us	mg_small  Name [CUDA memcp [CUDA memcp apply_kern cudaStream cudaStream cudaFree cudaFree cudaFree cudaGetDevi cudaGetDevi cudaGetDevi cudaEventS cudaEventS cudaEventC cudaEventC cudaEventC cudaEventC cudaEventC cudaDevice	yy HtoD] yy DtoH] cl_device. create calMem  Attribute csprc cnchronize cord Name cestroy apsedTime eate	.3x3(unsigi ) :ties							char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time (%) 37.13% 32.17% 38.70% unsigned c 94.39% 1.42% 0.65% 0.65% 0.12% 0.12% 0.10% 0.05% 0.05% 0.05%	Time 89.028us 77.121us 73.605us 73.605us 139.14ms 2.0861ms 1.5806ms 1.5806ms 1.5806ms 172.57us 460.25us 172.138us 461.731us 460.75us 472.17us 472.1	/a.out embo  Calls 28 21 7, int, floa 49 49 49 188 88 28 2 7 14 28 7 63	Avg 3.1779us 3.67720us 10.515us t*) 19.877ms 19.877ms 32.257us 28.259us 5.1660us 24.652us 46.720us 46.720us 46.724us 46.724us 46.724us 46.724us 46.724us 46.724us 5.878us 1.7890us 5.8930us 283ns 283ns 283ns 283ns	Min 704ns 3.6160us 10.369us 2.60s 2.2850us 2.46ns 1.0390ms 2.46ns 10.254us 4.60.25us 13.9080us 4.673us 9.0130us 1.2770us 1.2770us 1.2770us 5.6690us 2.46ns 2	139. 139. 1.04 214. 49.5 40.4 11.6 40.4 11.7 5.26 4.34 6.56	Max 480us 440us 440us 944us .07ms 471ms .44us .95us 531us .25us 004us 922us 723us 787us 060us 660ns	mg_small  Name [CUDA memc; CUDA memc; apply_kern  cudaStream  cudevicefot  cudaFree  cudaMalloc  cudaFree  cudaGetDeviceGet  cudaVeren  cudsVeren  cudsVer	yy HtoD] yy DtoH] cl_device. reate calMem  Attribute csync ccerroper cord Name lestroy capsedfime cate cynchroniz gument	.3x3(unsigi ) :ties							char∗,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resul Time (%) 37.13% 32.17% 30.70% unsigned c 94.39% 1.42% 0.66% 0.65% 0.61% 0.12% 0.12% 0.12% 0.12% 0.05% 0.05%	Time 89.028us 77.121us 73.608us 73.625us 73.608us 73.608us 73.608us 73.608us 73.608us 73.608us 73.608us 74.138us 74.138us 74.138us 74.14.191us 67.520us 64.9272us 41.251us 73.844us 73.8444us 73.8448us 73.8448us 73.8448us	/a.out embo  Calls 28 21 7, int, floa 188 49 49 17 7 28 28 28 28 28 7 14 63 7	Avg 3.1790us 10.515us t*) 19.877ms 32.257us 28.257us 28.257us 24.652us 60.25us 50.880us 46.25us 50.880us 45.748us 9.7290us 5.8930us 283ns 564ns 564ns	Min 704ns 3.6160us 1.6369us 9.2850us 1.0390ms 6.6210us 7.890us 2.840s 2.54us 4.60.25us 1.6250us 3.6530us 3.9860us 4.773us 9.0130us 4.5070us 1.2770us 2.46ns 367ns 367ns 367ns	139. 139. 1.04. 148. 201. 49.5 460. 40.4 11.6 46.7 11.7 5.26 4.34	Max 440us 4440us 9444us .07ms 471ms .44us .95us .95us 531us .45us 904us 922us 787us 440us 660us 660us	mg_small  Name [CUDA memc; [CUDA memc; paply_kenric  cudaStream  cudeviceTot  cudaMalloc  cudsFream  cudawiceGet  cudaMemc;  cudaLaunch  cudaLaunch  cudaEventS;  cudaEventS;  cudaEventS	yy HtoD] yy Dodd  yy Dodd  yy Dodd  yy Caele create calMem Attribute sync caeProper' nchronize cord Name cestroy apsedTime ceate synchroniz gument ireCall	.3x3(unsigi ) :ties							char*,
GPU ac	48== Profi 48== Profi Type ctivities: ed char*,	ling resulting r	Time 89.028us 77.121us 73.605us 73.605us 139.14ms 2.0861ms 1.5806ms 1.5806ms 1.5806ms 172.57us 460.25us 172.138us 461.731us 460.75us 472.17us 472.1	/a.out embo  Calls 28 21 7, int, floa 49 49 49 188 88 28 2 7 14 28 7 63	Avg 3.1779us 3.67720us 10.515us t*) 19.877ms 19.877ms 32.257us 28.259us 5.1660us 24.652us 46.720us 46.720us 46.724us 46.724us 46.724us 46.724us 46.724us 46.724us 5.878us 1.7890us 5.8930us 283ns 283ns 283ns 283ns	Min 794ns 3.6169us 1.0399us 6.6190us 2.46ns 1.0399us 2.46ns 1.0399us 4.60.25us 3.6539us 3.9880us 4.673us 5.6699us 1.2779us 5.6699us 2.46ns 367ns 367ns 367ns 356ns	5.22 3.72 10.5 139, 1.04 201. 449. 460. 461. 5.22 4.34 6.56	utput_i  Max 480us 440us 944us .07ms 471ms .96us .95us .531us .25us 448us 902us 723us 787us 060us 660ns 908ns	mg_small  Name [CUDA memc; CUDA memc; apply_kern  cudaStream  cudevicefot  cudaFree  cudaMalloc  cudaFree  cudaGetDeviceGet  cudaVeren  cudsVeren  cudsVer	y HtoD] yy DtoH] il-device. ireate alMem  Attribut sync cePropert cord nchroniz cord apsedTim eate gument ireCall Count	.3x3(unsigi ) :ties							char∗,

# Выводы

- Развертка циклов и использование shared-памяти помогает значительно(вплоть до 20%) ускорить вычисления на GPU.
- В случае, когда в программе требуется вызывать несколько раз обработку на GPU, можно достичь ускорение(вплоть до 30%) при помощи CUDA-потоков.
- Транспорт по-прежнему является узким местом данной технологии.