



Hmmer on GPU

Шульга Ольга
Михеев Максим, Biodatomics

Актуальность проблемы

- Один из основных методов в метогеномике является поиск гомологичных последовательностей для данных NGS
- Существуют версии алгоритмов обработки данных, полученных в результате NGS, работающих на GPU старого поколения [3]
- Ещё не было попыток разработки версии Hmmer, работающей на GPU

Цель

- Разработка open source версии Hmmer, которая будет работать на GPU и FPGA.

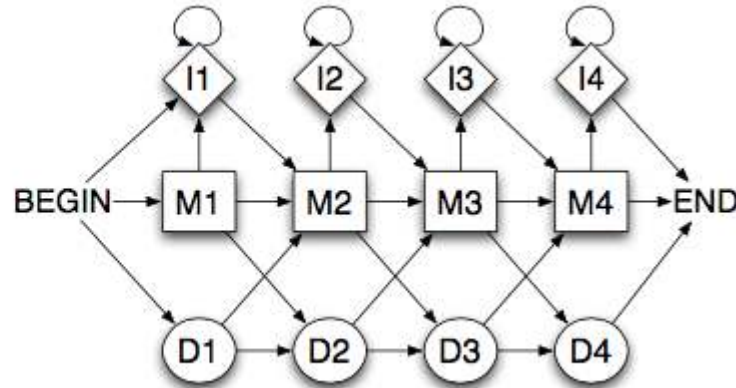
Задачи

- Изучение предметной области
- Выявление наиболее медленных модулей пакета Hmmer
- Адаптация алгоритма для работы на GPU
- Разработка прототипа версии Hmmer, работающей на GPU

HMMER

Hmmer - программа для поиска гомологичных последовательностей, работающая на основе скрытых марковских цепей (HMM - hidden Markov model).

s1	LE	---	VK
s2	LD	---	IR
s3	LE	---	IK
s4	LD	---	VE
query1	LDA	--	VK
query2	LDAA	VK	
query3	L	---	VK



NVIDIA GPU

- Имеет множество ядер. Программный код выполняется одновременно на всех ядрах.
- Если программа запустилась на GPU NVIDIA серии, например, GeForce или Quadro, то она запустится и на Tesla.
- Асинхронное исполнение функции ядра. Следует использовать средства синхронизации.

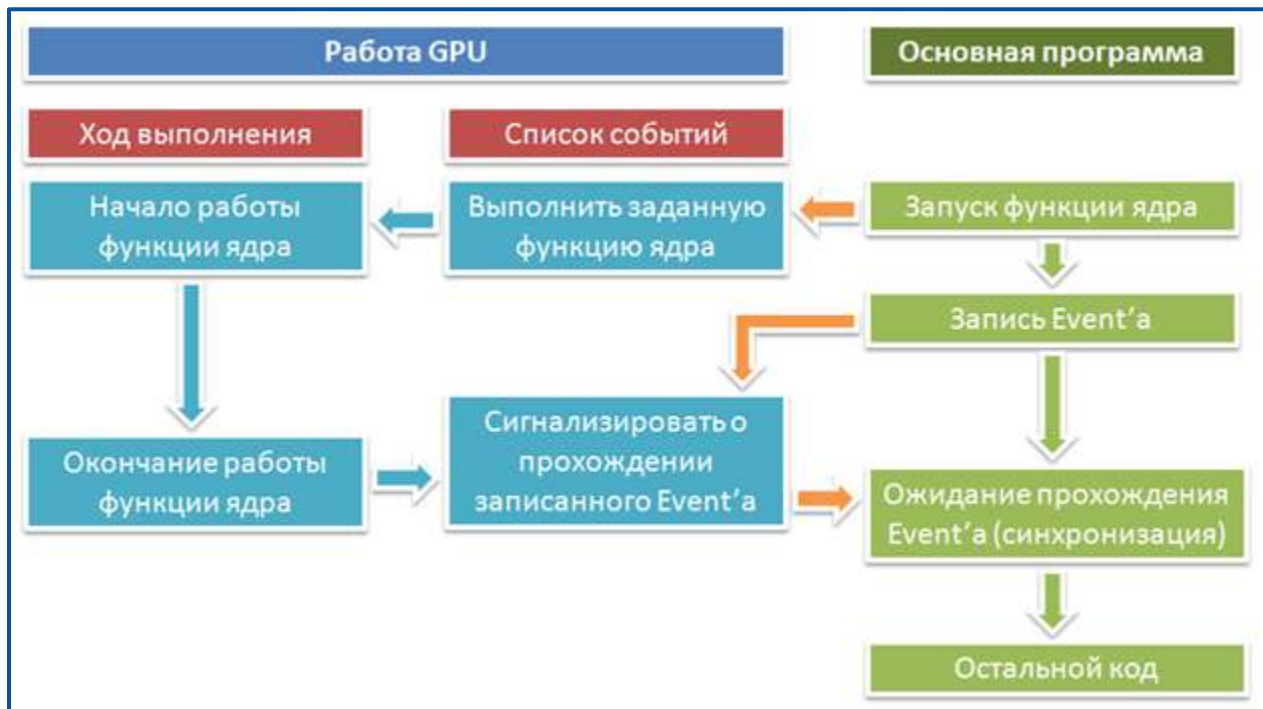
Tesla GPU от NVIDIA

Tesla GPU от NVIDIA - считается GPU нового поколения, используется для анализа больших данных.
[3]

SELECT THE RIGHT TESLA GPU

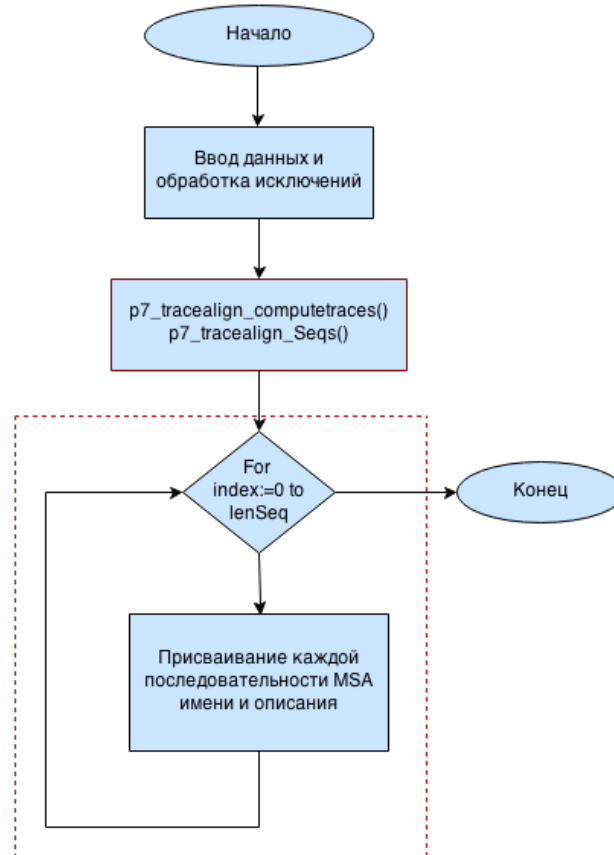
Features	Tesla K40	Tesla K20X	Tesla K20	Tesla K10
Number and Type of GPU	1 Kepler GK110B	1 Kepler GK110		2 Kepler GK104s
Peak double precision floating point performance	1.43 Tflops	1.31 Tflops	1.17 Tflops	0.19 Tflops
Peak single precision floating point performance	4.29 Tflops	3.95 Tflops	3.52 Tflops	4.58 Tflops
Memory bandwidth (ECC off)	288 GB/sec	250 GB/sec	208 GB/sec	320 GB/sec
Memory size (GDDR5)	12 GB	6 GB	5 GB	8 GB
CUDA cores	2880	2688	2496	2 x 1536

NVIDIA GPU

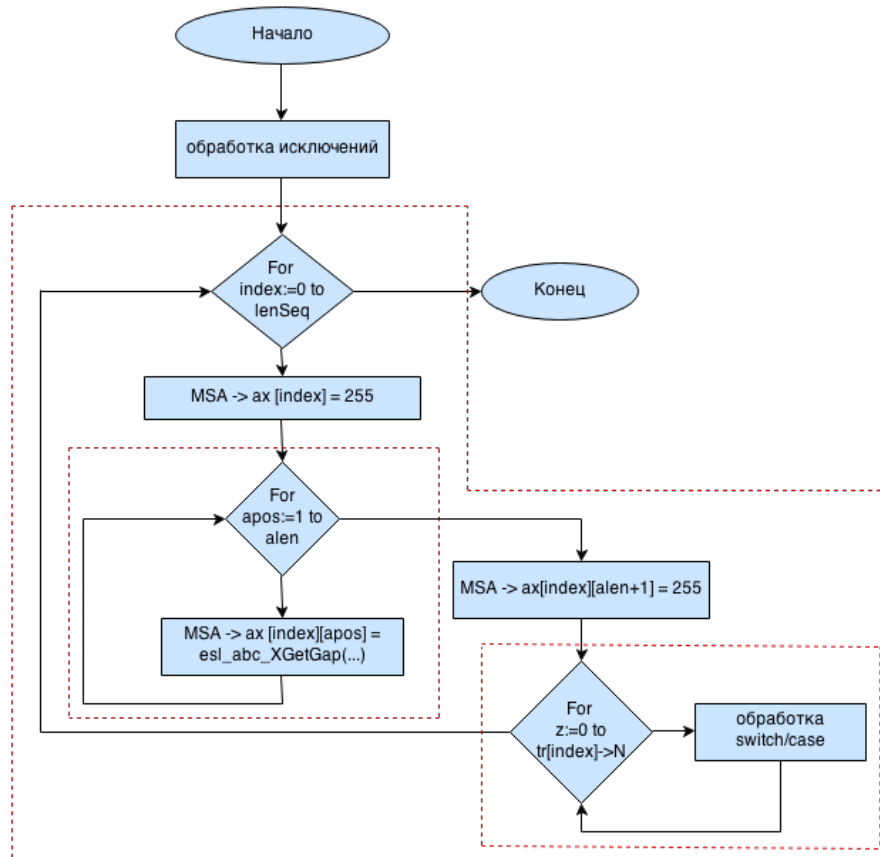


Hmmeralign

1) схема работы main()



Hmmeralign



2) схема работы
make_digital_msa()
(аналогично с
make_text_msa())

Ускорение алгоритмов с помощью GPU

- Smith-Waterman on GPU [2] - получилось ускорить алгоритм до 30x
- BLAST on GPU (Tera-BLAST) - производит анализ в 400 раз быстрее, чем изначальная версия (не open source)[3]
- Hmmer3 - распараллеливание некоторых частей алгоритма на CPU[4]

Задачи

- ~~Изучение предметной области~~
- ~~Выявление наиболее медленных модулей пакета Hmmer~~
- ~~Адаптация алгоритма для работы на CPU~~
- ~~Разработка прототипа версии Hmmer, работающей на GPU~~

Ссылки

1. <http://bioinformatics.oxfordjournals.org/content/early/2013/07/09/bioinformatics.btt403.full.pdf>
2. <http://www.biomedcentral.com/content/pdf/1471-2105-9-S2-S10.pdf>
3. http://www.timelogic.com/documents/TimeLogic_Tera-BLAST_whitepaper_v1.0.pdf
4. <http://selab.janelia.org/people/eddys/blog/?p=759>
5. и другие



Hmmer on GPU

Шульга Ольга
Михеев Максим, Biodatomics