CUDA-enabled implementation of whole-cell model

Hoang-Trong Minh Tuan
thoangtr@gmu.edu
Supervisor: Prof. Saleet Jafri

Department of BCB
George Mason University
Manassas, VA MSN 5B3, USA

Agenda

- Calcium handling in cardiac cells
- Calcium sparks: elementary events
- Challenge in whole-cell modeling
- CUDA programming model + Tesla(Fermi) architecture
- Benchmark results
- Conclusion + Future works

1. Calcium handling in cardiac cells
2. Calcium sparks: elementary events
3. Challenge in whole-cell modeling
4. CUDA programming model + Tesla(Fermi) architecture
5. Benchmark results
6. Conclusion + Future works

Review

- Sydney Ringer (papers: 1882, 1883): modern understanding of calcium ions in the contraction of the heart
- Bertil Hille, George Eisenman ... (1960-90s): detail understanding of ionic selectivity of ion channels
- Denis Noble, Silvio Weiderman ... (1950-current): full understanding of cardiac refractoriness (action potential)
- Alex Fabiato, Jon Lederer, Richard Tsien, Don Bers ... : more insights into the role of Ca^{2+} to excitation-contraction coupling mechanism (Calcium-induced Calcium-release, Ca^{2+}-spark, Ca^{2+}-waves)

Calcium ions

- Calcium ions do its functions via binding/unbinding mechanism
- Resting calcium concentrations: 1.8 mM extracellular, 1mM internal calcium storage SR, 1uM in cytosol
- Calcium elevation range from 10 to 100 fold in AP.
- Mechanism of calcium elevation: calcium-induced calcium-release (CICR)
- Elementary events of calcium elevation: Ca^{2+} sparks

Calcium handling

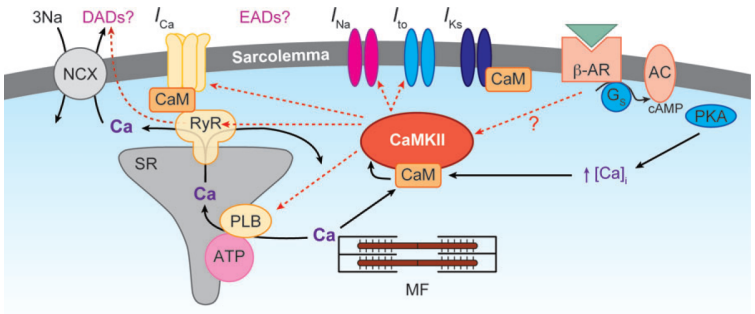


Figure: Ca^{2+} cycling

21st century: Understanding the molecular basic of calcium handling using data from gene expression to the structure and function of ion channels, regulatory proteins...

1 Calcium handling in cardiac cells
2 Calcium sparks: elementary events
3 Challenge in whole-cell modeling
4 CUDA programming model + Tesla(Fermi) architecture
5 Benchmark results
6 Conclusion + Future works

Spontaneous Ca^{2+} spark

Local-control theory

Global calcium elevation is the summation of discrete, elementary Ca^{2+} signalling events known as "spark".

Ca^{2+} spark

Cheng-Lederer-Cannell (1993) observed Ca^{2+} spark in quiescent rat ventricular myocyte.

Diadic supspace

The tiny space, composed by a cluster of LCC is in closed proximity to a cluster of RyR, at which Ca^{2+} occur.

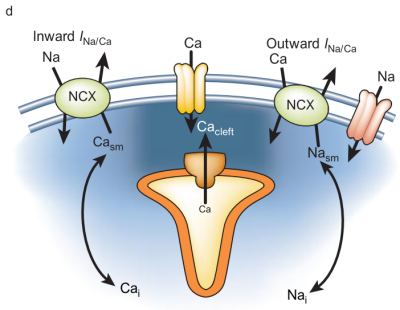


Figure: Junctional and submembrane space (Bers, 2002)

Diadic subspace

- 20,000 CRUs
- RyR:LCC varies upon species
- RyR opening rate 10^{-4} per second

Restoration of Ca^{2+} homeostasis:

- sequestering into ER/SR by SERCA2a pumps : 70%
- extrusion across plasma membrane (NCX, PMCA): 27%
- pumping to mitochondria via uniporter: 3%

1 Calcium handling in cardiac cells
2 Calcium sparks: elementary events
3 Challenge in whole-cell modeling
4 CUDA programming model + Tesla(Fermi) architecture
5 Benchmark results
6 Conclusion + Future works

Scientific questions

1. Mechanism of Ca^{2+} leaks
2. Mechanism of Ca^{2+} spark termination
3. Duration/Magnitude of a Ca^{2+} spark
4. How many LCC/RyR open in a spark?
5. How calcium elevation in one site ignite the calcium elevation in neighboring sites under normal vs. calcium overload?
6. ...

21st century: We need a detail temporospatial whole-cell model.

1 Calcium handling in cardiac cells
2 Calcium sparks: elementary events
3 Challenge in whole-cell modeling
4 CUDA programming model + Tesla(Fermi) architecture
5 Benchmark results
6 Conclusion + Future works

CPU vs. GPU

Cores

- CPU: 2, 4, 6, 8, 12 cores (Intel Westmere: 6-core)
- Tesla 1 (chip G80s): 128 "cores"
- Tesla 2 (chip GT200): 240 "cores"
- Fermi (chip GF100) : 448 "cores" (theory: 512)

Parallel threads

- CPU: each core can run 1 or 2 threads
- Tesla 1: maximum 12,288 "resident" threads
- Tesla 2: maximum 30,720 "resident" threads
- Fermi: maximum 21,504 "resident" threads

CPU vs. GPU

Registers

- CPU: dozens to a hundred of registers
- Tesla 1: 8K 32-bit registers
- Tesla 2: 16K 32-bit registers
- Fermi : 32K 32-bit registers

Shared & Cache Memory

- Intel Westmere: 32K L1D per core, 256KB L2 per core, 12MB L3 cache
- AMD Opteron Quad-core Istanbul: 64K L1D per core, 512KB L2 per core (Quad-core), 6MB L3 cache
- Tesla 1, Tesla 2: ???
- Fermi: per SM (64KB constant-cache memory, 64KB configurable L1-cache/shared memory), 768KB L2 cache.

CUDA Programming model

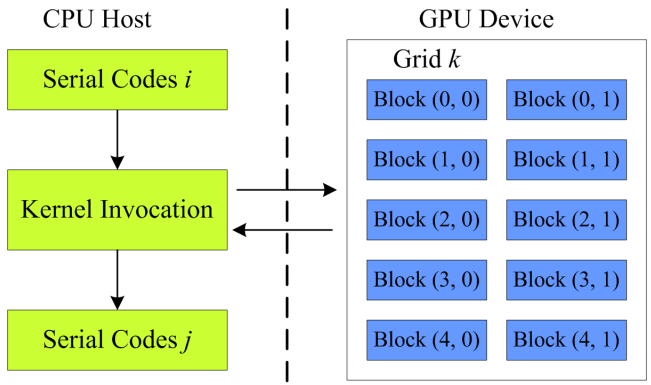


Figure: CPU-GPU code execution model

CUDA kernel invocation

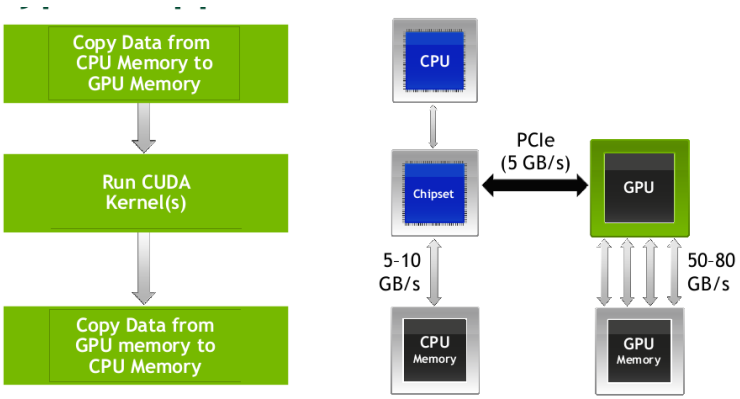


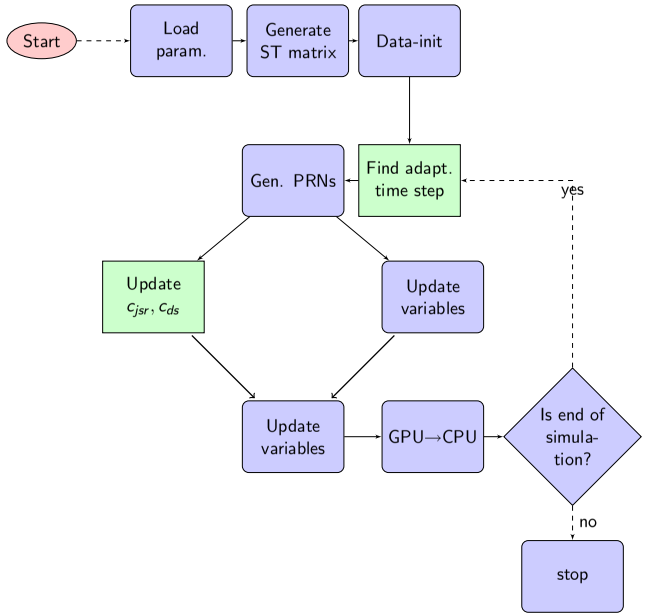
Figure: Kernel invocation + memory bandwidth

Good algorithm on GPU

1. Minimize data copy CPU-GPU
2. Minimize GPU global memory access
3. High temporal/spatial shared/cache memory "locality"
4. Reduce "registers" pressure

1 Calcium handling in cardiac cells
2 Calcium sparks: elementary events
3 Challenge in whole-cell modeling
4 CUDA programming model + Tesla(Fermi) architecture
5 Benchmark results
6 Conclusion + Future works

Flowchart



Benchmarking system

- Intel Quad-core Nehalem E5540 (8MB cache, 2.53 GHz)
- 6x4GB DDR3 1.3 GHz RAM
- nVIDIA Fermi C2050 GPU
- SATA2 1TB hard-drive (32MB buffer)

LEAK model

- 19,968 CRUs, LCCs are disabled (pharmacology disabled), each CRU has 50 2-state RyRs
- simulation 1 sec of physiological time, adaptive time step 10 us - 10 ns

System	No I/O	with I/O (single-time step)
CPU	20 min	21m38sec
CPU-GPU	2m7sec	8m47sec
Speed-up	10x	2.5x

Table: Benchmark LEAK model

Full ECC model

- 19,968 CRUs, each CRU has 50 2-state RyRs + 7 6-state LCCs.
- simulation 1 sec of physiological time, adaptive time step 10 us - 10 ns
- V-clamp 1Hz in 100ms

System	No I/O	with I/O (every 10 time-steps)
CPU	69m30sec	70m52sec
CPU-GPU	3m40sec	4m27sec
Speed-up	19x	16x

Table: Benchmark ECC model

1 Calcium handling in cardiac cells

2 Calcium sparks: elementary events

3 Challenge in whole-cell modeling

4 CUDA programming model + Tesla(Fermi) architecture

5 Benchmark results

6 Conclusion + Future works

Conclusion

- Heterogeneous computing in which CPU combines with GPU parallel processing power is the future trend.
- CUDA programming model and the new Fermi architecture is a powerful device, which has proved its effectiveness when being applied in computational cellular modeling
- Our novel method can be easily extended by incorporating other models easily
- Exciting and unaswered scientific questions can be potentially resolved using our method in a tractable time limit.

On-going research

- Building spatiotemporal 3D model of whole-cell rat ventricular myocytes
- Applying the model + method to studying different cellular pathological conditions.
- Extending the current whole-cell model (LCC, RyR, CAMKII, Calmodulin)
- Developing UI (dynamic plot)

Acknowledgements

- NSF, NIH
- nVIDIA
- Dr. Saleet Jafri (GMU, UMBC)
- Dr. George “Blair” Williams (GMU, UMBC)
- Dr. Jon Lederer (U.Maryland, BC)
- Steve Worley
- Rashmi Kumar