UNIVERSITY OF DELAWARE®

SC18

Dallas, TX | hpc inspires.

The University of Delaware is proud of its long participation history at the supercomputing conference. Our decades of continuous contribution to the area of High Performance Computing and Parallel Computing, which covers computer architectures, compiler technology, runtime technology, programming models and applications, is proof of on-going wonderful research and development at UD. We are committed to continuing our efforts towards building future software and hardware technologies and contributing to high quality research. Thank you for visiting our booth.

--The UD @ SC Team

Participants

Professors

Sunita Chandrasekaran (CIS)
Rudolf Eigenmann (ECE)
Guang R. Gao (ECE)
Juan Perilla (Chemistry)

Students

(CIS) Joshua Davis
(CIS) Sanhu Li
(CIS) Mauricio Ferrato
(CIS) Robert Searles
(Chemistry) Alex Bryer
Alex Bryer (Chemistry)
Akshay Bhosale (ECE)
Diego A. Roa (ECE)
Tan Xu (ECE)
Jose M Monsalve (ECE)

Booth posters

"Caviness Community Cluster"
J Frey et al (UD-IT)

"Energetics of nucleotide translocation through HIV-1 CA hexamer"
C Xu et al (Chemistry)

"**Inositol phosphates are assembly co-factors for HIV-1"**
C. Xu et al (Chemistry)

"Abstractions and Directives for Adapting Wavefront Algorithms to Future Architectures"
R. Searles et al (CIS)

"Accelerated Sequence Alignment algorithm for Next Generation Sequencing"
L. Sanhu et al (CIS)

"Landing Codelets PXM on DARTS. DEMAC - Towards a Hardware/ Software Codesign Evaluation Platform"
D. Roa and R. Kabrik et al (ECE)

"A Collaborative Defense Against Wear Out Attacks in Non-Volatile Processors"
P. Cronin et al (ECE)

"Power-and Endurance-Aware Neura Network Training in NVM-based Platforms"
F Meng et al (ECE)

"Compiler-directed Energy optimizations for STT-RAM based Scratchpad Memory"
F. Hosseini et al (ECE)

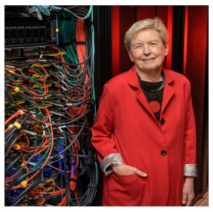
"Automatic Optimization of Computational Applications"
A. Bhosale et al (ECE)

"CI+PT: A CI+MBPT extension combining configuration interaction and Møller-Plesset perturbation theory for valence space"
C Cheung et al (Physics)

Caviness Community Cluster

J.Frey, A. Schwartz, S. Sunkara, and W. Totten • IT HPC Team
frey@udel.edu, anita@udel.edu, ssunkara@udel.edu, totten@udel.edu

UD's third HPC community cluster is named in honor of Jane Caviness, former director of Academic Computing Services at the University of Delaware, program director for NSFNET, deputy division director of NSF's Division of Networking and Communications Research and Infrastructure, and vice-president for Networking at EDUCOM.



Penguin Computing was chosen to build the *Caviness* cluster. Their innovative *Tundra* design is based on the **Open Compute Project (OCP)**, a standards-based initiative with the goal of making the hardware of computation more efficient, flexible, and scalable. OCP lets IT pack a diverse array of computational hardware in a smaller amount of space using less power than traditional clusters.

The *Caviness* cluster uses **Intel Xeon** processors. But two 18-core Broadwell CPUs make each node in *Caviness* about 2x more productive. An **Intel Omni-Path** 100 Gbps network links all the nodes, providing nearly 2x the communications bandwidth available in *Farber*.

The *Caviness* cluster uses the **Slurm Workload Manager** to schedule and execute work. As an open source project, IT can extend the functionality, debug and fix problems, and give back to the extensive community of Slurm users. Many of the top supercomputing sites also use Slurm, making *Caviness* users better prepared to work on other systems.

The *Caviness* cluster carries forward several technologies present in UD's Mills and Farber clusters:

- 200 TB of Lustre high-speed scratch storage
- 130 TB of home and workgroup storage
  - Daily snapshots backed-up off-site

However, as this cluster grows storage is added in each new rack. Thus, both types scale in capacity and Lustre also scales in performance.

Nearly all hardware in the first generation of the *Caviness* cluster has already been purchased by stakeholders. This fall, IT staff will be working on the specification and acquisition of a first addition of compute capacity to the cluster.

Node Specs (Generation 1)	Count	Available
2 x 18C Intel CPU, 128 GB RAM	60	1
2 x 18C Intel CPU, 256 GB RAM	48	6
2 x 18C Intel CPU, 512 GB RAM	6	0
2 x 18C Intel CPU, 2 x nVidia P100, 128 GB RAM	3	2
2 x 18C Intel CPU, 2 x nVidia P100, 256 GB RAM	5	1
2 x 18C Intel CPU, 2 x nVidia P100, 512 GB RAM	1	0

Visit our website to find out more about community clusters and how you can get involved!

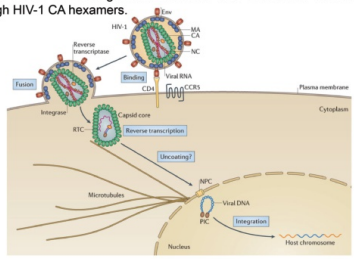
Energetics of nucleotide translocation through HIV-1 CA hexamer

Chaoyi Xu and Juan R. Perilla

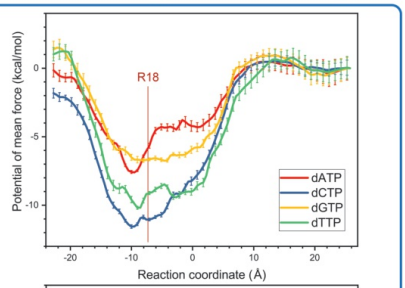
Department of Chemistry and Biochemistry, University of Delaware

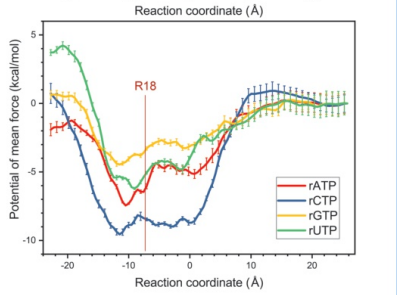
Early stage of HIV-1 infection

The capsid core encases the viral genome, in particular a mature HIV-1 capsid composes of hundreds of capsid hexamers and twelve capsid pentamers (CA). HIV-1 replication is highly dependent on its capsid [1]. During the early stage of HIV-1 infective cycle [1], the capsid traffics the viral genome towards the host nucleus. Somewhere along the way reverse transcription is triggered inside the capsid which results in capsid uncoating and releasing of the viral DNA. Recently, a study of the physical properties of a whole HIV-1 casid established its ion semi-permeability [2]. Furthermore, recent experimental studies indicate that the center pore of a CA hexamer is a channel that regulates dNTP translocation required for reverse transcription [3]. Nonetheless, the molecular mechanism of dNTPs translocation through a CA pore is still unclear. Here, we investigate the translocation of dNTPs and rNTPs using Hamiltonian replica exchange molecular dynamics simulations for a total sampling over 100 microseconds. Altogether, our results reveal the energetics associated with nucleotide translocation through HIV-1 CA hexamers.

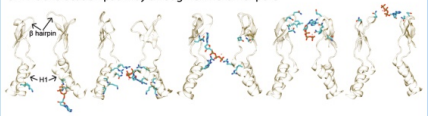


Gibbs free energy of NTP binding

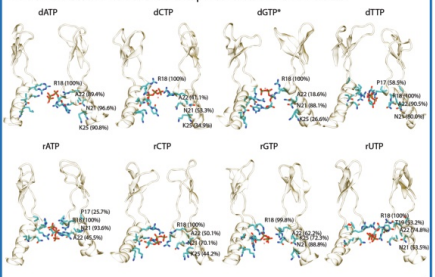




dATP translocation pathway through CA hexamer pore

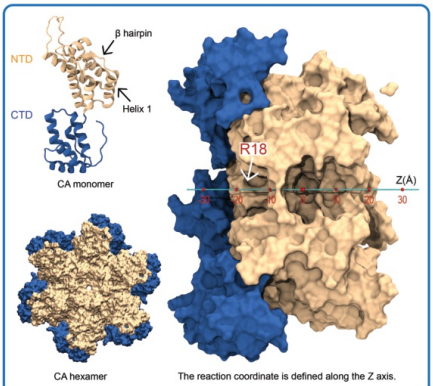


NTP-CA molecular interactions and top 4 contacts at their bind sites

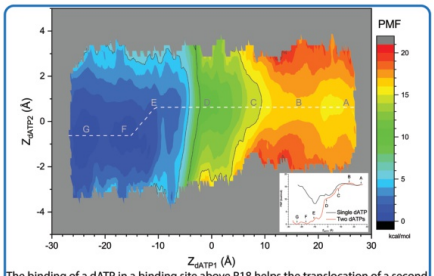


* The PMF of dGTP translocation does not have an evident minimum but the dGTP-CA interaction at around -10 Å is shown as a comparison.

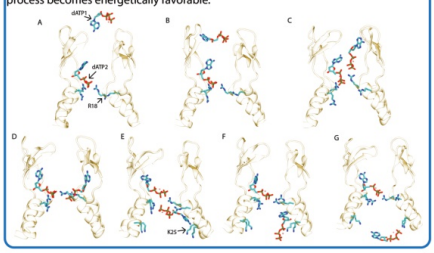
Architecture of the dNTP channel



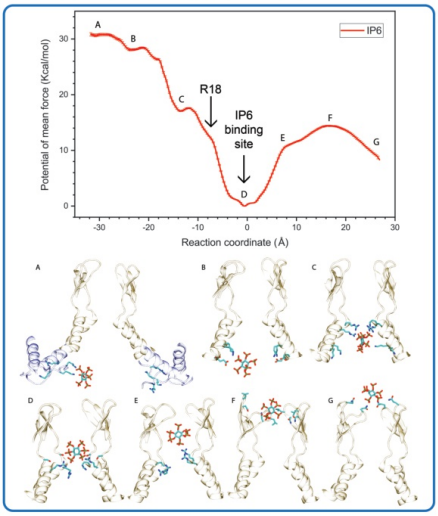
Cooperative translocation of dATPs



The binding of a dATP in a binding site above R18 helps the translocation of a second incoming dATP from CA hexamer exterior to interior, making the dATP translocation process becomes energetically favorable.

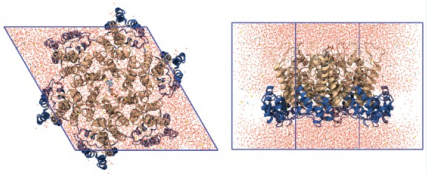


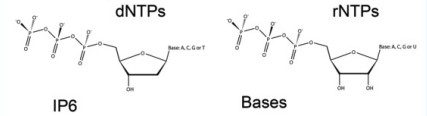
Gibbs free energy of IP6 binding

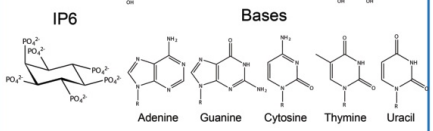


Methods

A capsid hexamer [4] embedded in a P6 latice

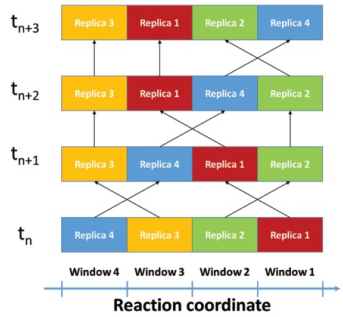






All-atom molecular dynamics simulations

All-atom simulations were performed using the v-RESPA algorithm in NAMD2.12 [5] employing the CHARMM36m force field [6]. Temperature and pressure were held constant at 310 K and 1 atm, respectively. Simulations were performed at 150mM NaCl and 50mM Ca. Gibbs free energies were calculated along the reaction coordinate via potential of mean force calculations (PMF). PMFs were calculated using the Hamiltonian replica exchange method (HREX) [7-8]. HREX enables the swapping between neighboring umbrella sampling windows resulting in improved sampling efficiency. The initial coordinates for each HREX simulation were obtained from steered-MD along the reaction coordinate and equilibrated for 20 ns per window. A force constant of 2.5 Kcal/mol Å^2 was applied to restrain the NTP/IP6 molecule in each window. Exchange attempts between neighborin gwindows were made every 2 picoseconds, employing a metropolis-hasting exchange rule.



Conclusions

► Single NTP binding is highly energetically favorable (-10 kcal/mol) in the channel and is mediated by residues R18 and K25. Therefore, a passive translocation of single nucleotides seems unlikely.

► Distinct CA-nucleotide contacts give rise to different binding profiles. In particular, rGTP exhibits lower binding affinity compared to other nucleotide types.

► By introducing a dATP binds above R18, the translocation of an incoming dATP becomes energetically favorable (-15 Kcal/mol). A cooperative dNTP translocation mechanism is proposed.

► The binding site of IP6 locates above R18 in the channel.

References

[1] Campbell, Edward M., and Thomas J. Hope. HIV-1 capsid: the multifaceted key player in HIV-1 infection. Nature Reviews Microbiology (2015): 471.
[2] Perilla, Juan R., and Klaus Schulten. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. Nature communications (2017): 15959.
[3] Jacques, David A., et al. HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. Nature 536.7616 (2016): 349.
[4] Gres, Anna T., et al. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. Science (2015): aaa5936.
[5] Phillips, James C., et al. Scalable molecular dynamics with NAMD. Journal of computational chemistry (2005): 1781-1802.
[6] Jing Huang et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods (2016): 71-73
[7] Jiang, Wei, et al. Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD. Computer physics communications (2014): 908-916.
[8] Ostmeyer, Jared, et al. "Recovery from slow inactivation in K+ channels is controlled by water molecules." Nature (2013): 121.

This work is supported by NIH grants P30GM110758-04 and P50GM082251-11.
Computational resources were provided by NCSA Blue Waters and TACC Stampede2.

Inositol phosphates are assembly co-factors for HIV-1

Chaoyi Xu[1], Robert A. Dick[2], Kaneil K. Zadrozny[3], Florian K. M. Schur4,5, Jonathan M.Wagner[3], Barbie K Ganser-Pornillos[3], Owen Pornillos[3], Volker M. Vogt[2], Juan R. Perilla[1]*

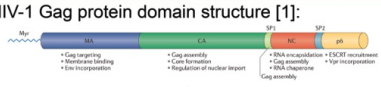
[1]Department of Chemistry and Biochemistry, University of Delaware, [2]Department of Molecular Biology and Genetics, Cornell University, [3]Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine. [4]Structural and Computational Biology Unit, EMBL Heidelberg, [5]Institute of Science and Technology Austria.

Abstract

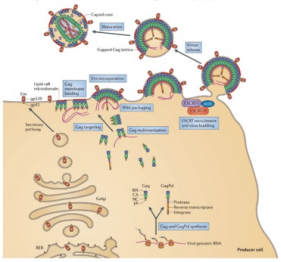
During the late stage of the HIV-1 life cycle, the HIV-1 Gag proteins aggregate at the plasma membrane. As gradually being cleaved by the viral proteases, Gag protein assembles into two structurally different lattices, namely immature and mature assembly, respectively. The Gag assembly and maturation is triggered by the conformational changes of a 14 amino acid peptide called SP1. Here, we show that a group of negatively-charged small molecules, called inositol phosphates (IPs), including IP3, IP4, IP5 and IP6, engage in the SP1 folding and promote both immature and mature Gag lattice formation. These small molecules are presented in all mammalian cells. We therefore identified IPs as natural cellular co-factors which manipulate HIV-1 assembly and maturation.

HIV-1 Gag assembly

HIV-1 Gag protein domain structure [1]:

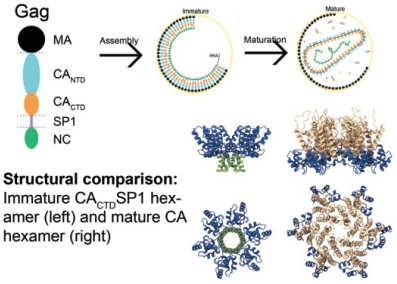


Gag assembly in HIV-1 life cycle [1]:



SP1 is a molecule switch [2-4] of Gag protein assembly in the late stage of HIV-1 life cycle.

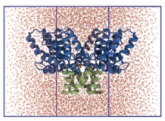
1. During virus assembly, it triggers the formation of spherical virus-like particles (VLPs) underneath the membrane.
2. During maturation, CA-SP1 is cleaved by the viral protease, which frees the immature CA lattice and leads to assembly of the mature conical capsid.



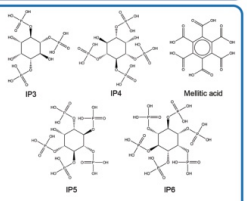
Structural comparison: Immature CA$_{CTD}$SP1 hexamer (left) and mature CA hexamer (right)

Methods

Model building:



CA$_{CTD}$SP1 in a P6 water box

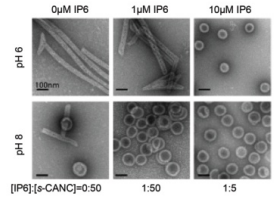


MD simulation:

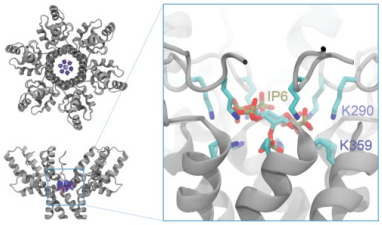
1. Force fields: CHARMM 36m [5], CHARMM general force field [6], TIP3P water [7].
2. All models underwent minimization, thermalization and equilibration before the 2 μs production runs.
3. NPT ensemble was used, 310 K and 1 atm.
4. IP6 bound CA$_{CTD}$SP1 model was simulated on PSC Anton2 [8], while the rest of the model were simulated on TACC Stampede2 using NAMD2.12 [9].
5. IP6 translocation PMF was calculated by Umbrella sampling/Hamiltonian replica exchange method [10].

IP6 stabilizes HIV-1 immature lattice

IP6 was found to be able to dramatically promote CANC assembles into immature VLPs in vitro.

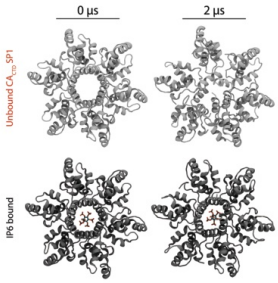


IP6 molecule is found in the immature X-ray crystal structure, bound with LYS290 and LYS359.

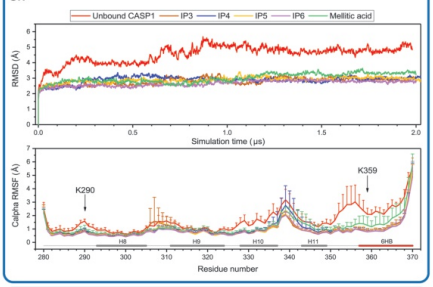


How can IP6 promote CANC to form the immature lattice?

MD simulation results demonstrate that IP6 can stabilize CASP1 immature lattice, especially the six-helix bundle (6HB).



The RMSD and RMSF data indicate the additional flexibility in the unbound CASP1, particularly the 6HB region. Moreover, this stabilizing effect is not specific to IP6, other IPs and mellitic acid can also reduce the structural flexibility of CASP1 hexamer.

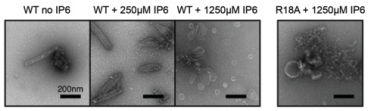


References

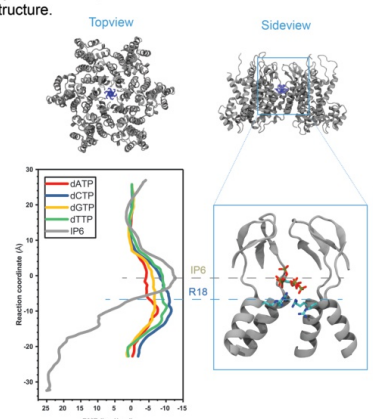
[1] Freed, Eric O. "HIV-1 assembly, release and maturation." Nature Reviews Microbiology 13.8 (2015): 484.
[2] Gross, Ingolf, et al. "A conformational switch controlling HIV-1 morphogenesis." The EMBO journal 19.1 (2000): 103-113.
[3] Schur, Florian KM, et al. "An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation." Science 353.6298 (2016): 506-508.
[4] Wagner, Jonathan M., et al. "Crystal structure of an HIV assembly and maturation switch." Elife 5 (2016).
[5] Huang, Jing, et al. "CHARMM36m: an improved force field for folded and intrinsically disordered proteins." nature methods 14.1 (2017): 71.
[6] Vanommeslaeghe, Kenno, et al. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields." Journal of computational chemistry 31.4 (2010): 671-690.
[7] Jorgensen, William L., and Corky Jenson. "Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density." Journal of Computational Chemistry 19.10 (1998): 1179-1186.
[8] Shaw, David E., et al. "Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer." Proceedings of the international conference for high performance computing, networking, storage and analysis. IEEE Press, 2014.
[9] Phillips, James C., et al. "Scalable molecular dynamics with NAMD." Journal of computational chemistry 26.16 (2005): 1781-1802.
[10] Jiang, Wei, et al. "Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer." Journal of chemical theory and computation 8.11 (2012): 4672-4680.
[11] Murphy, R. Elliot, et al. "Solution Structure and Membrane Interaction of the Cytoplasmic Tail of HIV-1 gp41 Protein." Structure 25.11 (2017): 1708-1718.
[12] Datta, Siddhartha AK, et al. "Interactions between HIV-1 Gag molecules in solution: an inositol phosphate-mediated switch." Journal of molecular biology 365.3 (2007): 799-811.

Small molecules bind to mature CA

In addition, our experiment also found that CA is able to assemble into mature structures in the presence of IP6.



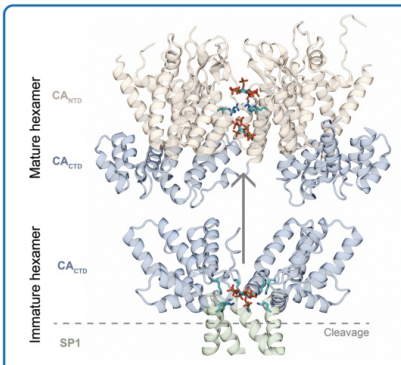
IP6 molecule forms salt bridges with R18 in a CA mature crystal structure, as with LYS290 and LYS359 in immature structure.



The PMF of IP6 translocation through the center channel of the mature CA hexamer suggests that:

1. the binding site of IP6 in the channel locates slightly above the R18 ring, which is consistent with the experimental data.
2. the translocation of IP6 into the channel from N-terminal domain is energetically favorable.

Conclusions



Our experimental and computational results lead to following conclusions:

1. IP6 facilitates formation of the six-helix CA-SP1 bundle by binding to the two rings of lysine residues at K290 and K359 at the center of the Gag immature hexamer.
2. During virus maturation, proteolytic removal of the SP1 segment disrupts the 6HB, thus releasing IP6 and at the same time unmasking the R18 binding site in mature CA. Subsequently, IP6 binds to the newly exposed R18 site in CA, promoting formation of CA hexamers and in turn the mature CA lattice.
3. IPs are assembly co-factors of HIV-1.

Future directions:

The assembly of HIV-1 Gag proteins at the plasma membrane (PM) is also driven by the viral Env incorporation. Early studies show that the Env incorporation is mediated by interactions between the Gag MA domain and the C-terminal cytoplasmic tail of Env gp41 (GP41CT). Recently, the solution structure of GP41CT and its interactions with PM has been resolved [11]. However, it is still unclear how the MA-GP41CT interacts and how these interactions influence Env incorporation and Gag assembly. These research questions will be investigated in our future study.

On the other hand, we wouldl also like to test how IPs interact with MA domain of Gap protein. Since it has been reported [12] that IP6 is able to interact with both ends of Gag protein, MA and NC domains, and induce its conformational changes.

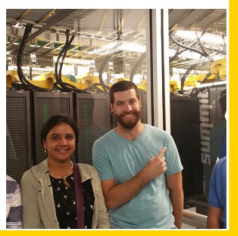
This work is supported by NIH grants P30GM110758-04 and P50GM082251-11. Computational resources were provided by PSC Anton2 and TACC Stampede2.

Abstractions and Directives for Adapting Wavefront Algorithms to Future Architectures

Robert Searles, Advisor: Sunita Chandrasekaran

Motivation

- Parallel programming is software's future
  - Acceleration
  - Programmability
- State-of-the-art abstractions handle simple parallel patterns well
- Complex parallel patterns are hard!

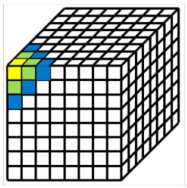


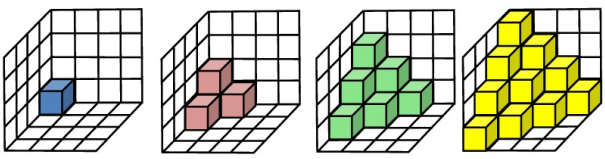
Contributions

- An abstract representation elucidating architectural, memory and threading challenges to programming models for such complex wavefront algorithms as used in Minisweep that can be broadly applicable to applications with a similar computational motif.
- A performance-portable implementation of these abstractions using OpenACC to offload portions of an application to a variety of parallel architectures found in modern HPC systems.
- A description of the challenges in existing programming models, and extensions that will allow programmers to over- come the obstacle of recursivity in the spatial dimensions of wavefront algorithms without requiring large modifications to the code base.

Abstractions for Wavefront Parallel Pattern

- Minisweep, a miniapp modeling the wavefront sweep component of ORNL's Denovo Radiation Transport code
- 80-99% of Denovo's runtime
- Early application readiness on Titan
- Used for Summit acceptance testing
- No existing solution **or software abstractions** in high-level programming models like OpenMP/OpenACC
- Created a prototype for the wavefront sweep kernel of Minisweep, applied Koch-Baker-Alcouffe (KBA), a parallel sweep algorithm that overcomes some of its dependencies
- Evaluated prototype on state-of-the-art HPC architectures, including NVIDIA P100 and V100 GPUs, ORNL's Titan Supercomputer, and ORNL's Summitdev cluster
- Examined scalability of proposed abstraction using MPI to run across nodes and GPUs within those nodes

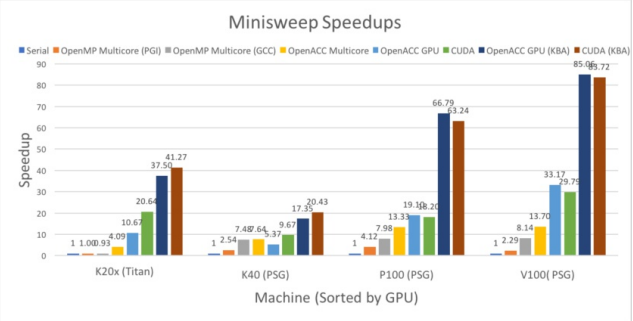




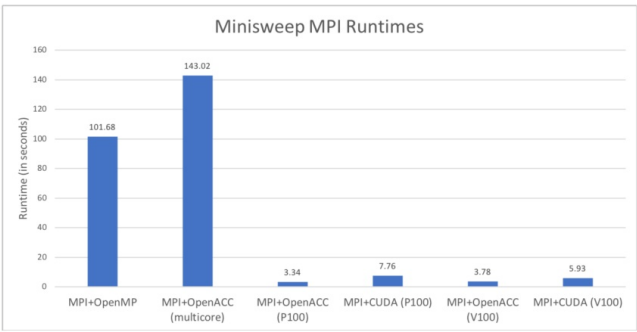
Performance Results

Machine	CPU	GPU
NVIDIA PSG (V100)	Intel Xeon E5-2698 (2x, 16 cores each)	NVIDIA Tesla V100 (32GB HBM2)
NVIDIA PSG (P100)	Intel Xeon E5-2698 (2x, 16 cores each)	NVIDIA Tesla P100 (16GB HBM2)
NVIDIA PSG (K40)	Intel Xeon E5-2690 (2x, 10 cores each)	NVIDIA Tesla K40 (12GB GDDR5)
ORNL Titan	AMD Opteron 6274 (16 cores)	NVIDIA Tesla K20X (6GB GDDR5)
ORNL Summitdev	IBM POWER8 (2x, 10 cores each)	NVIDIA Tesla P100 (16GB HBM2)

- PGI 18.4 Community Edition Compiler



- Single node/single GPU (higher is better)
- 85.06x speedup on NVIDIA Tesla V100



- 4 nodes/4 GPUs per node - 16 GPUs total
- MPI + OpenACC: 7.56x speedup compared to single GPU

Summary

- Proposed prototype for a high-level language extension
  - Improves programmer productivity and application portability
- Demonstrated prototype's applicability on a real-world scientific application
- Evaluated prototype on leadership-class supercomputers
- Github: https://github.com/rsearles35/minisweep

R. Searles, S. Chandrasekaran, W. Joubert, O. Hernandez, *"Abstractions and Directives for Adapting Wavefront Algorithms to Future Architectures,"* at The Platform for Advanced Scientific Computing (PASC) 2018. DOI: 10.1145/3218176.3218228

Future Work

- Investigate additional wavefront algorithms
- Build benchmark suite for wavefront applications
- Prototype abstraction
  - Extension to OpenACC
  - Evaluate on other wavefront algorithms

This material is based upon work supported by the National Science Foundation under grant no. 1814609.
We gratefully acknowledge the support of NVIDIA PSG Cluster for providing access to the P100 and V100 GPUs used for this research, as well as donating P40 and V100 GPUs to our group.

Accelerated Sequence Alignment algorithm for Next Generation Sequencing

Sanhu Li, Sunita Chandrasekaran

Motivation


PacBio Sequel


SOLiD 5500xl W


Ion Proton I


Illumina NovaSeq 6000 v4


Roche 454 GS FLX+ / FLX

Next Generation Sequencing Instruments are generating more data!

- Massive data generated from Next Generation Sequencing (NGS) instruments. 35 peta-bases per year currently, and the number is still increasing
- Different types of computation resource available, but can hardly be used

Research Problem

- Design and implement a faster sequence alignment algorithm that can make use of different types of computational resources without sacrificing accuracy, and is friendly to long sequences for future sequencing technology

State of the Art

Sequence Alignment Tool	HPC Platform	Year
BarraCUDA	CUDA and POSIX Threads	2012
Bowtie2	Intel Thread Building Blocks	2009
BWA	Multi-core CPU systems	2009
BWA-PSSM	Same with BWA	2014
NextGenMap	CUDA/OpenCL/POSIX Threads	2013
SparkBWA	Hadoop	2016
Subread	POSIX Threads	2016

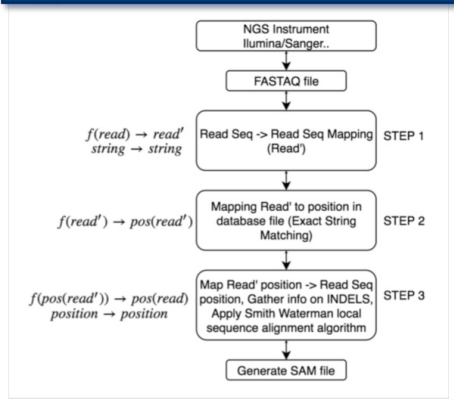
Challenges

- Inputs of sequence alignment algorithms are not exact sequences from the genome. The queries are usually DNA sequences with mutations and gaps.
- Current solutions are either hash based or FM-index based algorithms. Hash based algorithms are facing memory limits and are hard to implement on different types of accelerators. FM-index based algorithms reply FM-index algorithm which is slower than hash algorithms.
- Future sequencing instruments are more in favor of producing long reads and current tools are not really good at handling them.

Contributions

- Design and implement a new sequence alignment algorithm
- Design a new algorithm that can improve alignment performance of Fast Minute full text Index (FM-index)
- Accelerate our algorithm targeting multiple HPC architectures (Multicore CPU, Manycore GPU, etc.)

Approach



- We take a sliding window from the queries as the seeds in step 1. Seed length will be varied for different spicies.
- We use our novel LC-hash algorithm to map the seeds to locations in the genome database in step 2.
- We analyze the alignment locations of the seeds from step 2, and finally decide the location of the query. We use a customed version of edit distance algorithm called edlib to get better performance compared to the smit—waterman algorithm which is usually used in other tools.

Evaluation and Conclusion

We evaluate 1 million 100bp queries, 0.1 million 1kbp queries and 0.01 million 10kbp queries against yeast genome using our tool and BWA on a laptop with Intel i7-7700HQ. Both tools are running in serial mode.

We can see that AccSeq has a better serial performance compared to BWA-MEM, with a performance decrease for longer sequences, but is still faster than BWA in those cases.

Dataset	AccSeq	BWA-MEM
1M 100bp	66.58s	146.48s
0.1M 1kbp	101.47s	538.189s
0.01M 10kbp	364.73s	541.390s

Landing Codelets PXM on DARTS

DEMAC - Towards a Hardware/Software Codesign/Evaluation Platform

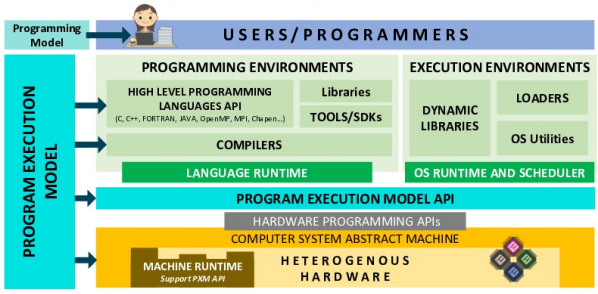
Diego Roa, Jose Monsalve, Ryan Kabrick, Xu Tan, Guang Gao

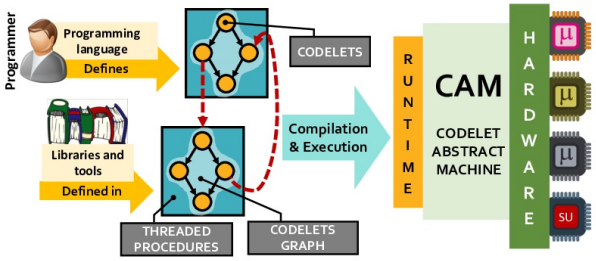
University of Delaware — CAPSL Computer Architecture and Parallel Systems Laboratory

SC18 Dallas, TX

Codeletes (PXM)

Program Execution Models (PXM) define the structure and behavior of a program when executed on an abstract target machine. For parallel systems PXMs describe the program in terms of **synchronization**, **memory models** and **actors** (**tasks**). A common PXM for parallel systems is a key element for the success of parallel programming, as it allows software modularity and portability across multiple systems.

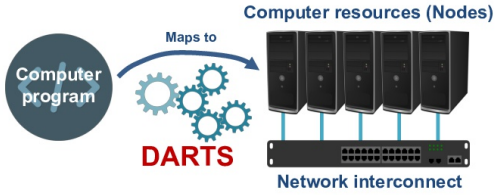


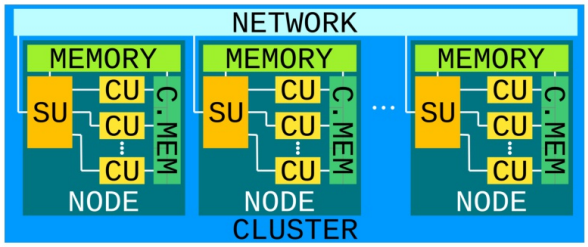


The Codelet model defines a program as a collection of Event and Data driven **tasks**, connected by **Dependencies,** forming a **Codelet Graph (CDG)**. **Threaded Procedures (TPs)** are asynchronous procedures defined by a CDG and its environment and required resources. TP communicate through continuations. Codelets borrow the semantics of Von Neumann PXMs, while CDGs are defined using Dataflow semantics.

DARTS

The Delaware Adaptive Runtime System (DARTS) is an implementation of the Codelet PXM that includes a set of **APIs** and **Compiler Tools** to define **Threaded Procedure** and **Codelets** as well as the mapping of the CAM to the Computer System. A set of libraries define the runtime that implements the behavior of the Codelet PXM.

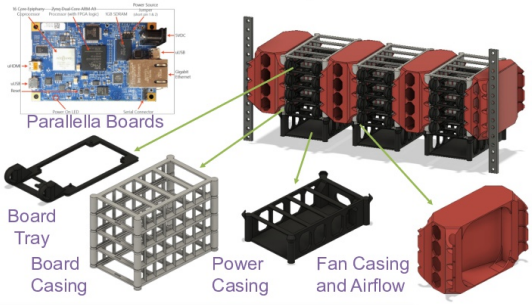




A **Codelet Abstract Machine (CAM)** is used as an abstraction of the underlying hardware. It is made out of **Computational Units** (codelet execution) and **Scheduling Units** (handles resources and synchronization). Memory can be placed at each level of the hierarchy.

DEMAC

The Delaware Modular Assembly Cluster (DEMAC) is an array of Parallella Embedded Systems that combines the many cores Epiphany chip and the embedded FPGA with the flexibility of a complete open source stack. The mount is house made 3D-printed frames allowing low cost implementation and scalability. It is design to fit 4Us of a standard size rack. Files for the rack design are open source. The multiple nodes allows us to explore distributed version of the Codelet Model where there is no notion of shared memory.



Developing hardware level support for the Codelet Model is possible through the use of the FPGA. The many RISC Cores of the Epiphany act as CUs, and the lack of cache coherency, allows us to explore more flexible memory models as well as different mapping mechanisms for the Codelet Abstract Machine. The low cost of the system and its open source promotes interdisciplinary collaborations and expansion of parallel computing to other fields (e.g. Industrial Automation and Robotics)

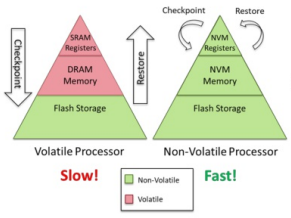
Project supported by NSF 1439165

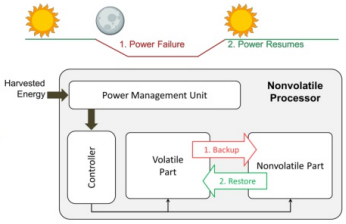
A Collaborative Defense Against Wear Out Attacks in Non-Volatile Processors

Patrick Cronin, Chengmo Yang, and Yongpan Liu

1. Introduction

In uncertain power situations, volatile processors must checkpoint and restore their registers and volatile RAM on every shutdown/restart. NVPs only checkpoint and restore their volatile registers, leading to fast operation.

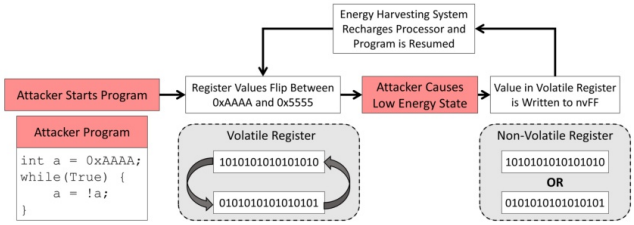




Non-volatile processors (NVPs) represent the next generation of Internet of Things (IoT) devices. NVPs are designed for intermittent power situations, utilizing a capacitor and energy harvesting for power delivery.

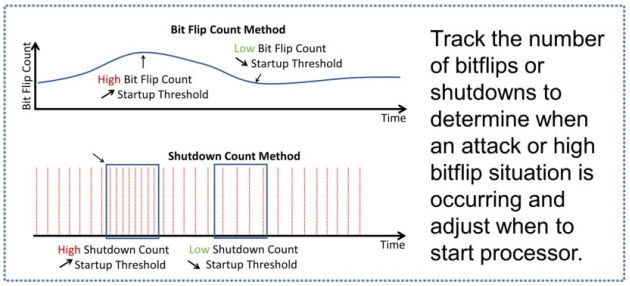
2. Problems!

Non-volatile memory experiences endurance issues. In an attack, a device could be worn out in a few days.

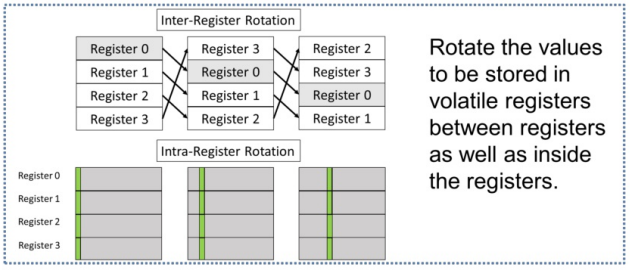


If an attacker can rapidly manipulate power or cause energy intensive situations they can cause rapid shutdowns, accelerating the NVPs demise. Our scheme detects these rapid shutdowns and applies mitigations.

3. Collaborative Wear Leveling



Track the number of bitflips or shutdowns to determine when an attack or high bitflip situation is occurring and adjust when to start processor.

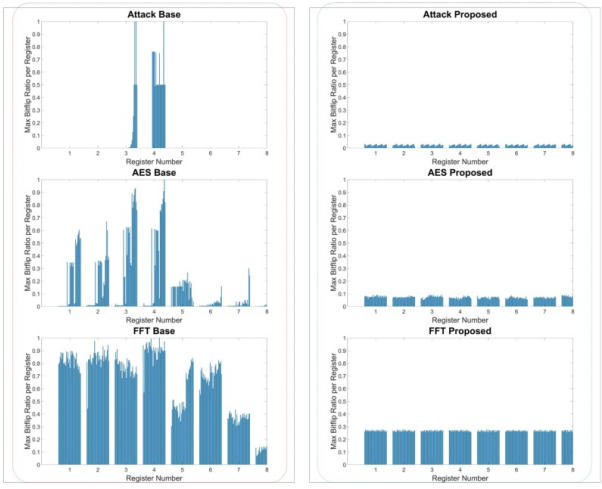


Rotate the values to be stored in volatile registers between registers as well as inside the registers.

Combine both startup manipulation and register rotation to greatly increase the lifetime of registers. Decrease the startup threshold when good behavior is detected in order to increase device responsiveness.

4. Results

Test system in GEM5 architecture simulator, running 1 attack kernel and 2 common kernels for embedded systems (AES and FFT).



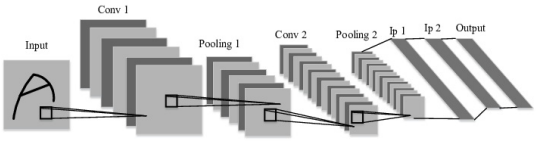
Thanks to NSF CNS-1739390 for the support

Power-and Endurance-Aware Neural Network Training in NVM-based Platforms

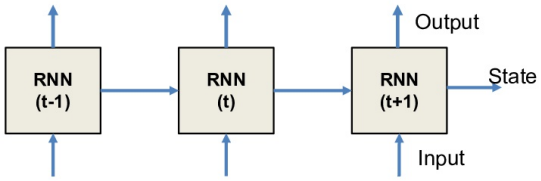
Fanruo Meng, Yuan Xue, Chengmo Yang

Introduction

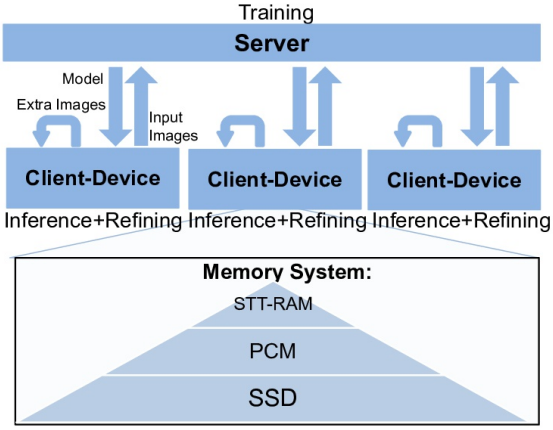
CNN



RNN



Training



	Non-volatile	Cell Size(F^2)	Read Latency(ns)	Write Latency(ns)	Leakage Power	Write Energy	Write Endurance
SRAM	No	120~200	Fast	Low	High	Low	10^{16}
DRAM	No	12-30	Medium	Medium	Medium	Medium	10^{16}
STT-RAM	Yes	20	Fast	High	Low	High	$>10^{12}$
PCM	Yes	4	Medium	High	Low	High	10^8~10^9

Our Goal

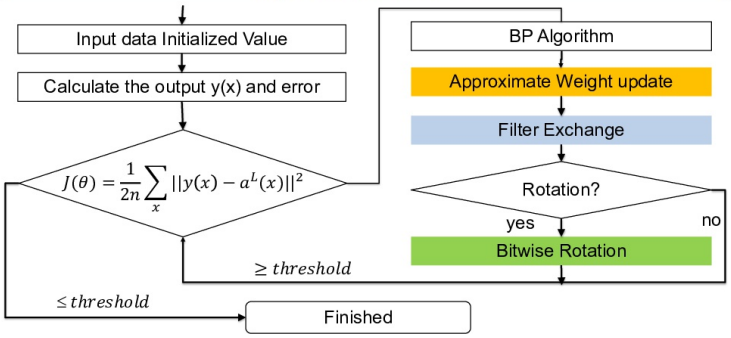
NVM Pros:

- ✓ High Data Density
- ✓ Low Leakage Power
- ✓ High Speed

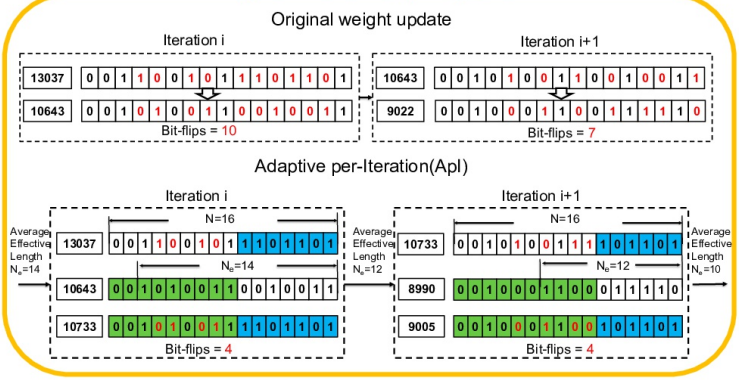
NVM Cons:

- × High Write Latency and Energy → Approximate Weight update
- × Limited Write Endurance → Filter Exchange, Bitwise Rotation

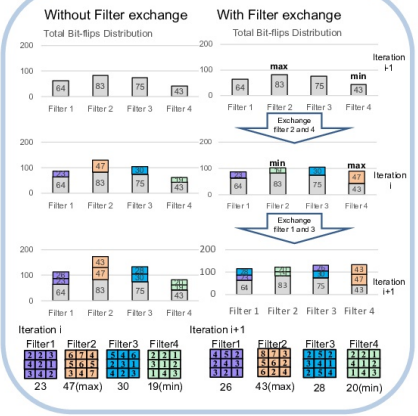
Algorithm Process



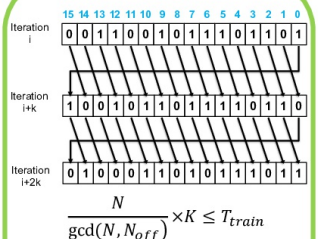
Approximate Weight update



Filter Exchange



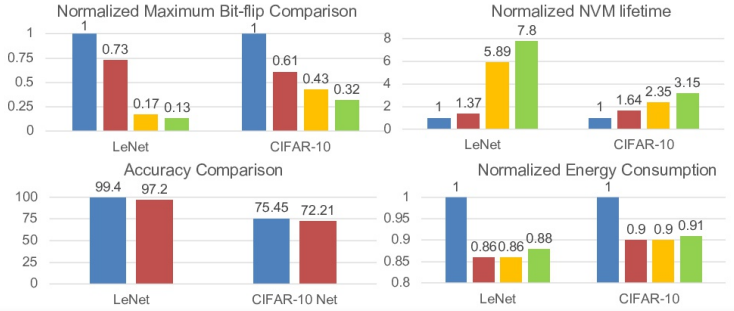
Bitwise Rotation



$$\frac{N}{\gcd(N, N_{off})} \times K \leq T_{train}$$

Variable	Explanation	Value
T_{train}	Total Training Iterations	10,000
N	Bit Length	16
N_{off}	Rotation Offset	9
K	Iteration Stride	625

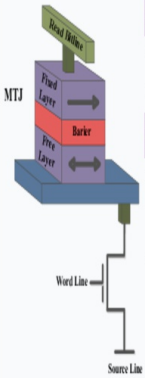
Results



Compiler-directed Energy optimizations for STT-RAM based Scratchpad Memory

Fateme Hosseini, Chengmo Yang
Electrical and Computer Engineering University of Delaware, Newark Delaware

1. STT-RAM based Scratchpad Memory



Scratchpad memory

- High-speed internal memory
- Software controlled

STT-RAM

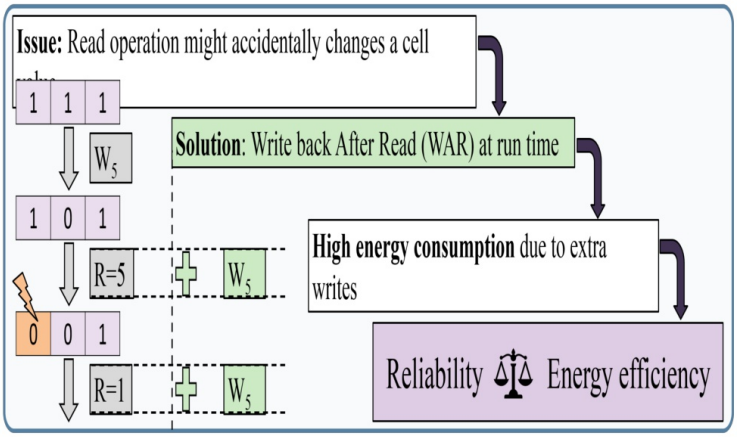
- Magnetic Tunnel Junction (MTJ) Cell

Write: Pulsing current **Read**: Detecting resistance

- Promising non-volatile memory technology

✓Scalability ✓Endurance
✓Performance ✓Low power leakage
✗Write energy ✗Read disturbance

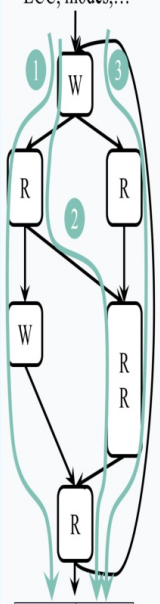
2. Read Disturbance Problem



3. Proposed Technique

Leveraging Compiler

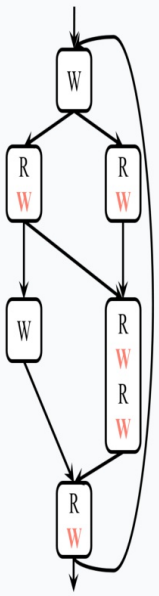
- Reliability- & energy-aware transformations
- No system structure overhead like parity, ECC, modes,...



Path	#W
1	2
2	1
3	1

(1)Explicit WAR

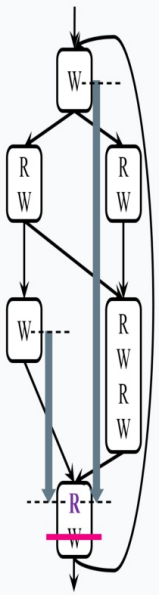
- WARs are added to code at compile time
- Eliminated and move to optimize the code



Path	#W	Cmp ORG
1	4	↑2X
2	5	↑5X
3	5	↑5X

(2) Dead reads

No WAR after last-loads where memory value won't be live anymore

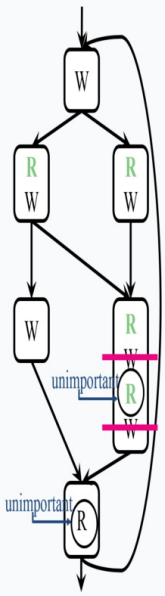


Path	#W	Cmp (1)
1	3	↓25%
2	4	↓20%
3	4	↓20%

(3) Read importance

Find hot-loads and evaluate

- No WAR for predecessors of an unimportant load

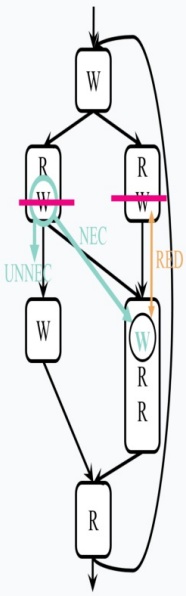


Path	#W	Cmp (2)
1	3	↓0%
2	2	↓50%
3	2	↓50%

(4) Selective WAR

Path-based analysis

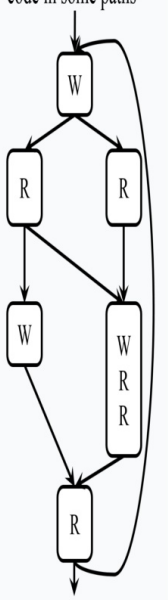
- Remove unnecessary WAR
- Merge redundant WAR



Path	#W	Cmp (3)
1	2	↓33%
2	2	↓0%
3	2	↓0%

Results

- 56% less energy compared to WAR
- No energy overhead compared to original code in some paths

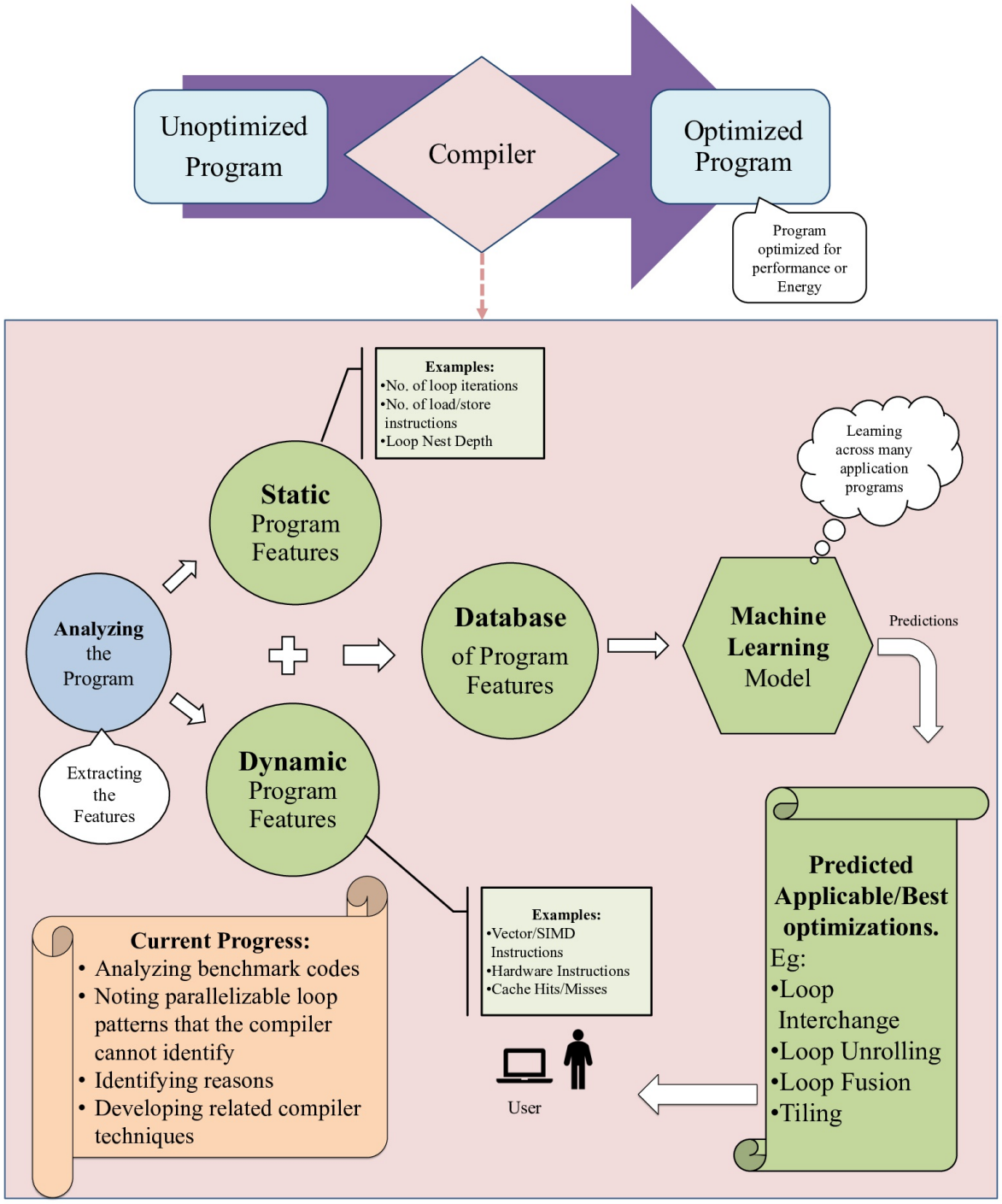


Path	#W	Cmp WAR
1	2	↓50%
2	2	↓60%
3	2	↓60%

Automatic Optimization of Computational Applications

Akshay Bhosale and Rudolf Eigenmann
Department of Electrical and Computer Engineering

OpenMP



CI+PT: A CI+MBPT extension combining configuration interaction and Møller-Plesset perturbation theory for valence space

C. Cheung[1], M. S. Safronova[1,2], S. G. Porsev[1,3], M. G. Kozlov[3,4], I. I. Tupitsyn[5]

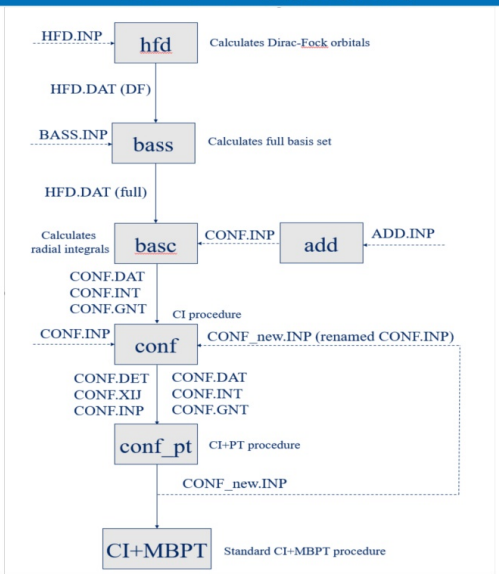
Abstract

The CI+MBPT package allows one to carry out relativistic atomic calculations for many-electron atomic systems, used for the prediction of atomic energy levels and other observables in the framework of relativistic quantum mechanics [1]. The configuration interaction method is used for the interaction between valence electrons, while many-body perturbation theory is used to calculate core-core and core-valence correlations. These two methods are combined to acquire benefits from both approaches and attain better accuracy. This is effective for atoms with 2-3 valence electrons when the valence CI space is not very large.

In this extension of the CI+MBPT package, CI+PT improves and optimizes the CI valence space by using CI in a small subspace and calculating 2nd order Møller-Plesset (MP2) corrections for the complementary subspace. We start with a small CI space and calculate 2nd order MP corrections for all states of interest. At the same time, weights of configurations are calculated in the 1st order correction to wavefunctions. In the next iteration, all configurations with weights above some threshold are incorporated in a new CI subspace with the complementary subspace under MP2 corrections. This can be repeated for an optimal valence CI space. After the valence CI space is formed, MBPT corrections can be added to account for core-valence correlations in the standard manner for CI+MBPT.

[1] M. G. Kozlov, S. G. Porsev, M. S. Safronova, and I. I. Tupitsyn, Computer Physics Communications **195**, 199 (2015), ISSN 0010-4655

CI+PT scheme



Description of programs

The *hfd* program solves restricted Hartree-Fock-Dirac equations self-consistently and finds two-component Dirac-Fock (DF) orbitals and eigenvalues of the HFD Hamiltonian. In this method, the orbitals are found in the central field approximation and depend only on principle, orbital, and total angular momentum quantum numbers *n*, *l*, and *j* [1].

The *bass* program constructs the full basis set by forming DF orbitals for the core and valence shells, then adding virtual orbitals to account for correlations. A reasonable basis set should consist of the orbitals mainly localized at the same distances from the origin as the valence orbitals [1].

The *add* program constructs a list of configurations defining the CI space by exciting electrons from a reference configuration to a set number of 'active' nonrelativistic shells [1].

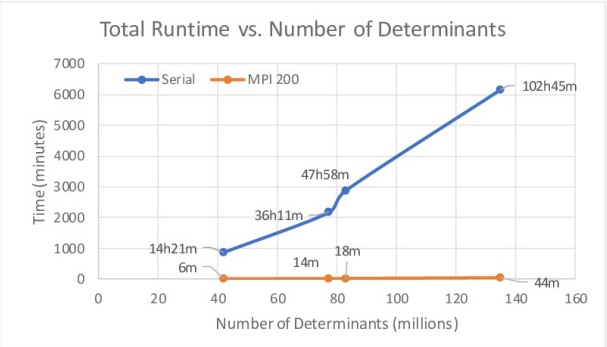
The *basc* program calculates one-electron and two-electron radial integrals, which are used by the program *conf* to form the CI Hamiltonian matrix in the CI space. The one-electron radial integrals correspond to the HF potential of the core, and the two-electron radial integrals account for the Coulomb and Breit interaction between valence electrons. [1].

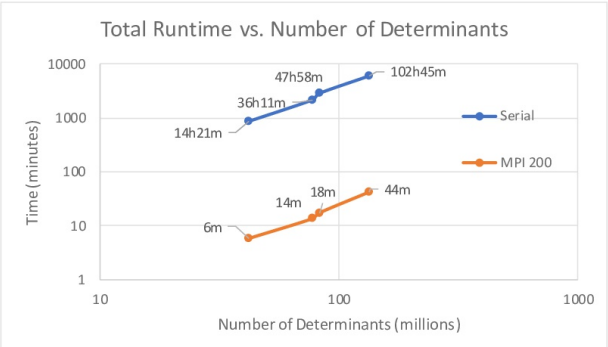
The *conf* program forms the Hamiltonian matrix and the matrix of the operator $\boldsymbol{J}^2$ in the configurational space of Slater determinants. Then the matrix eigenvalue equation is solved, and eigenvectors are saved to the file CONF.XIJ and eigenvalues are written to a final table in the file CONF.RES [1].

The *conf_pt* program calculates second order corrections to the energies and first order eigenvectors, and then calculates the weights of configurations and reorders the configuration list in order of descending weights. The MP2 energy corrections are written to a final table in CONF_PT.RES along with the original CI energy and total energies of the atomic energy levels.

MPI Parallelization of CI+PT code

The time consuming part of this code is *conf* and *conf_pt*. Depending on the size of the CI or PT space, the time it takes to run these codes are on the order of hours to days, depending on the number of determinants present in the calculation. In both *conf* and *conf_pt*, the matrix calculations are split into a diagonal calculation and an off-diagonal calculation. In order to decrease run times, an initial attempt has been made to parallelize the *conf_pt* code (parallelization of *conf* is in the works) using MPI. Below are the speedups of MPI with 200 cores on the University of Delaware's new *caviness* cluster.





[1]Department of Physics and Astronomy, University of Delaware, DE, USA
[2]Joint Quantum Institute, NIST and the University of Maryland, College Park, MD, USA
[3]Petersburg Nuclear Physics Institute, Gatchina 188300, Russia
[4]St. Petersburg Electrotechnical University "LETI", St. Petersburg, Russia
[5]Department of Physics, St. Petersburg State University, Ulianovskaya 1, Petrodvorets, St. Petersburg, 198504, Russia

This work was supported by U.S. NSF Grant No. PHY-1620687.

Booklet design by Jose Monsalve