Optimizations in CAD-based Monte Carlo Radiation Transport

Patrick C. Shriwise

PhD Defense

University of Wisconsin - Madison

June 8, 2018

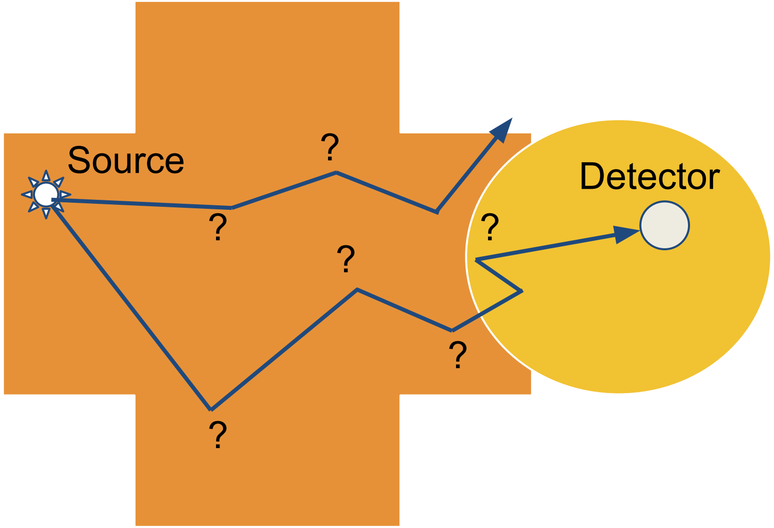
Outline

- Motivation
- Background
- Methods
  - Signed Distance Field Preconditioning
  - Ray Tracing Accelerations
  - Mesh Feature Adaptations
- Conclusion

Motivation

Monte Carlo Simulation

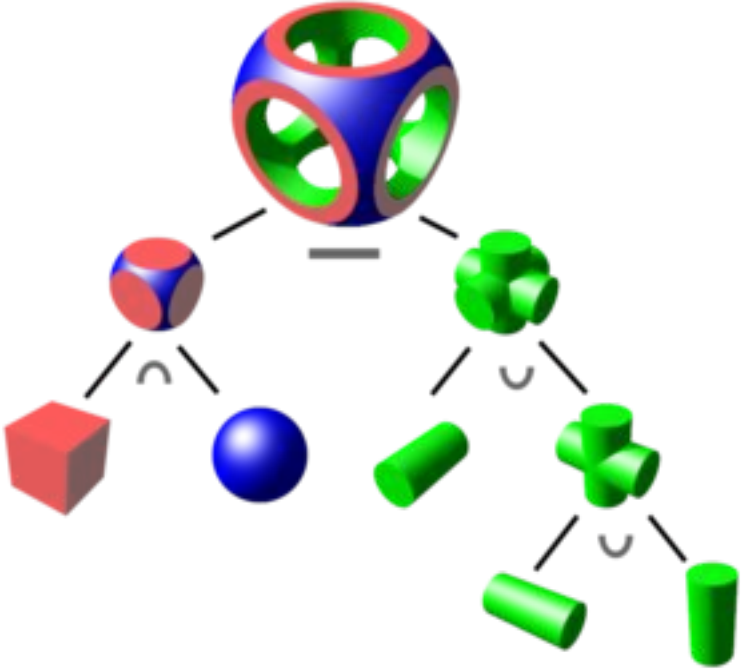
Treat each particle independently, randomly sampling probability distributions representing physical processes to statistically approach a solution to the system.



Constructive Solid Geometry (CSG)

Boolean combinations of analytic geometric objects

- Sphere
- Cylinder
- Parallelepiped
- Torus
- etc.

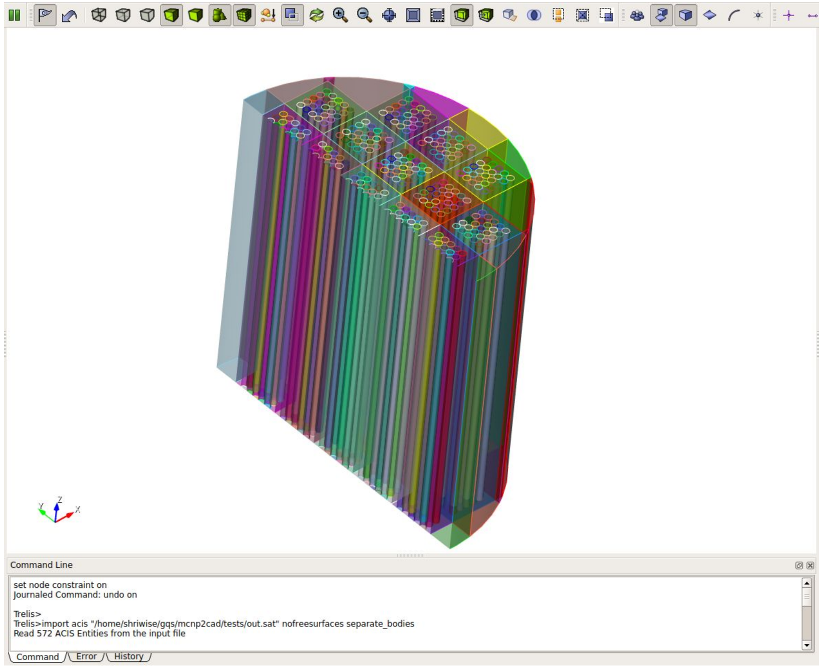


Computer Aided Design (CAD)

Text-Base CSG Input File

```
c example of pwrlat
      1   0 -1 -19 29 fill=1 imp:n=1
      2   2 -1 -301 302 -303 304 lat=1 u=1 imp:n=1 fill=-3:3
   -3:3 0:0
          1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2 2 2 2 2 1 1 2 2 2 2 2 1
          1 2 2 2 2 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1
      3   1 -19 -10 u=2 imp:n=1
      4   2 -1 #3 #5 #6 #7 #8 #9 #10 #11 #12 #13 #14 #15
   #16 #17 #18
          #19 #20 #21 #22 #23 #24 #25 #26 #27 #28 imp:n=1
   u=2
      5   like 3 but trcl=(-6 6 0)
      6   like 3 but trcl=(-3 6 0)
      7   like 3 but trcl=(0 6 0)
      8   like 3 but trcl=(3 6 0)
      9   like 3 but trcl=(6 6 0)
     10   like 3 but trcl=(-6 3 0)
```

Same Model in CAD Software



Why CAD-based geometry?

Human Factors

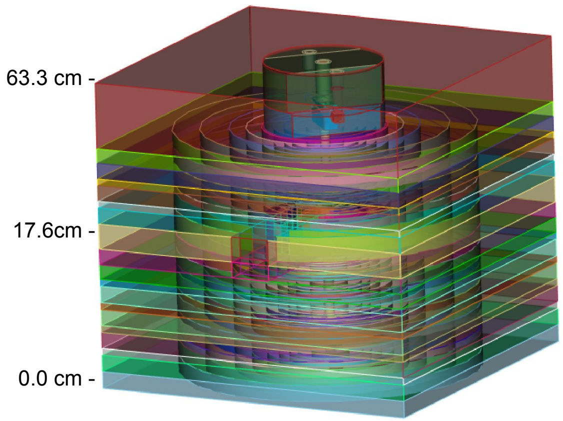
- interactive visualization and verification
- reduced cognitive load
- reduced human error

Technical Factors

- representation of higher-order surfaces
  - splines, Bezier curves, etc.
- shared domain for multi-physics simulations
  - structural, CFD, HT

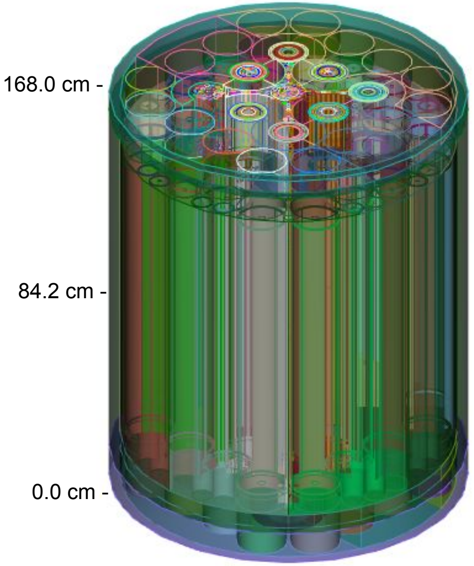
Benchmark Models

Frascati Neutron Generator
(FNG)



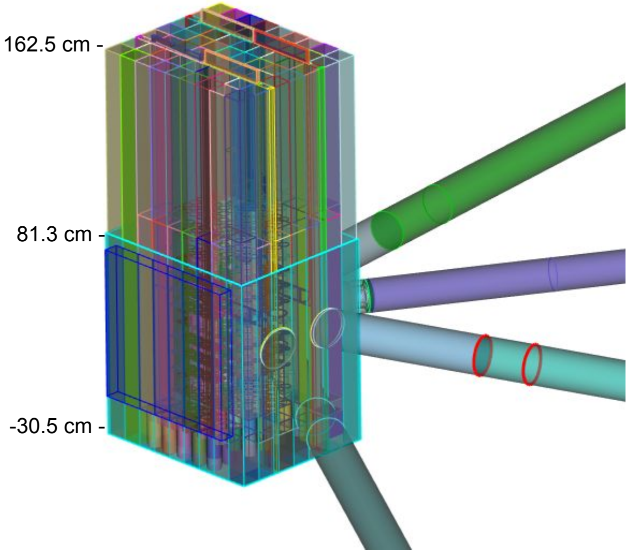
n, γ

Advanced Test Reactor
(ATR)



n

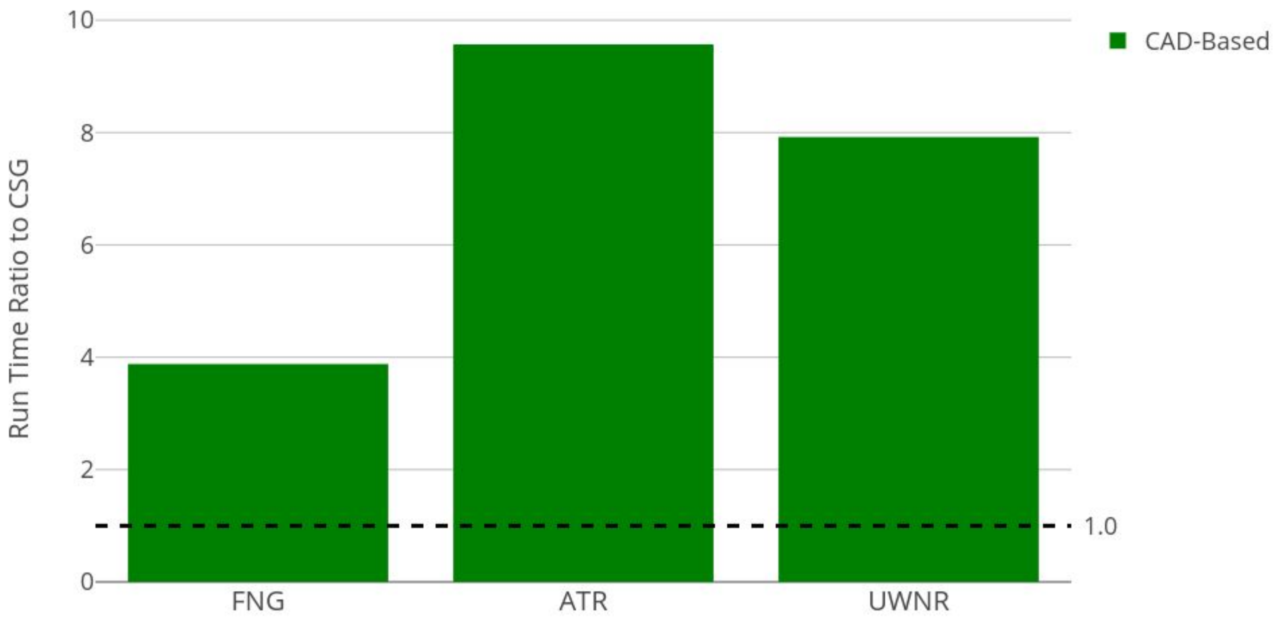
UW Nuclear Reactor
(UWNR)



n

Current State of Performance

CAD-based is anywhere from 2.5-10x slower than native code. Production analysis can take days/weeks.



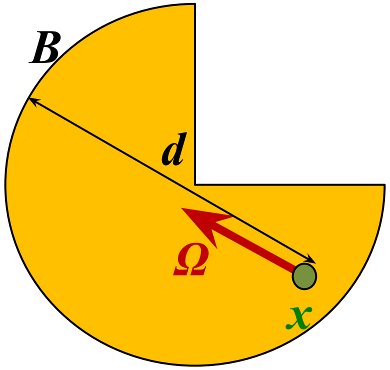
Problem Statement

Achieve CAD-based simulation times competitive with native CSG geometry representations with no change in CAD-based simulation results.

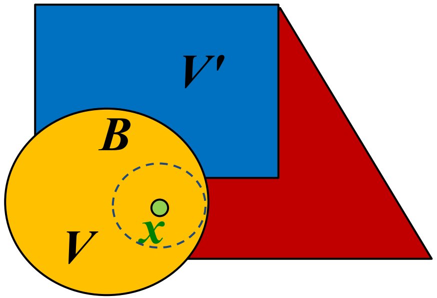
Background

Direct Accelerated Geometry Monte Carlo

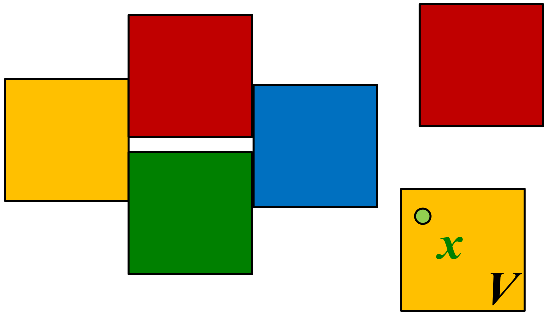
The Direct Accelerated Geometry Monte Carlo (DAGMC)[1] toolkit is used to track particles through discretized CAD geometries.



Distance to Boundary



Nearest Surface



Point Location

DAGMC Workflow

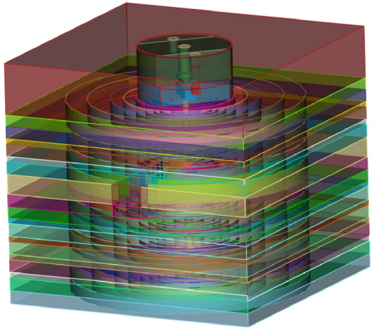
Simulation

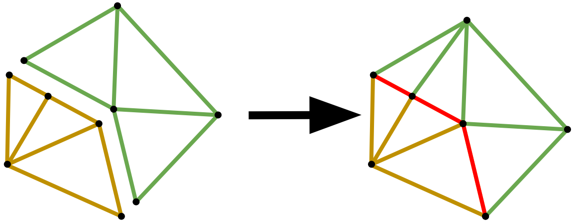
CAD (CUBIT/Trelis[2]) → Tessellation (MOAB[3]) → [DAGMC ↔ Physics]

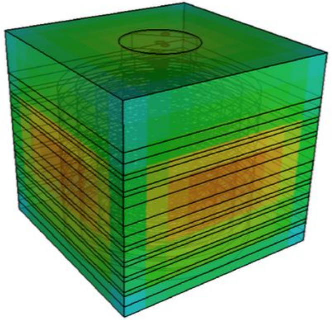
Tessellation: model sealing[4]

DAGMC: particle tracking

Physics: MCNP[5]

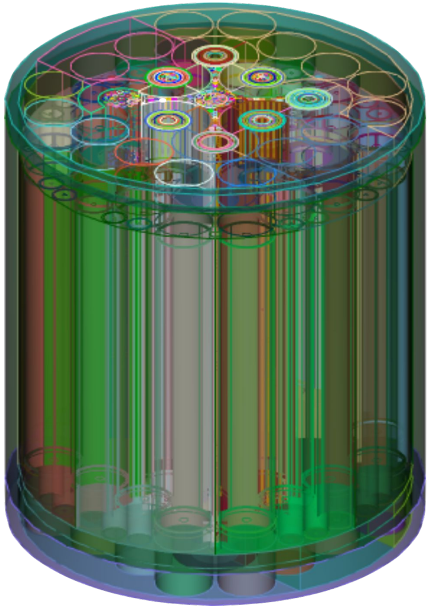


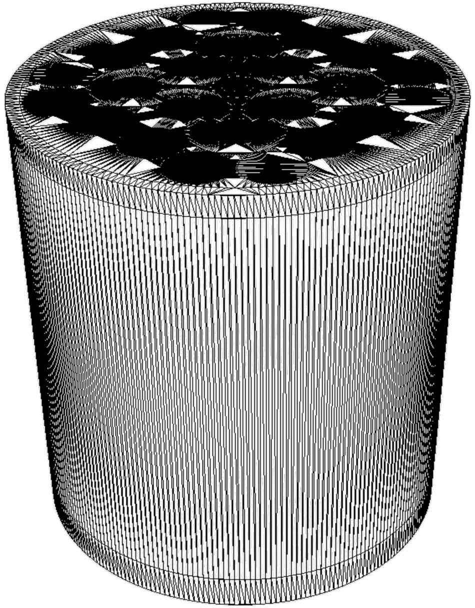




Tessellation

Advanced Test Reactor





Surface Mesh
(4.9M triangles)

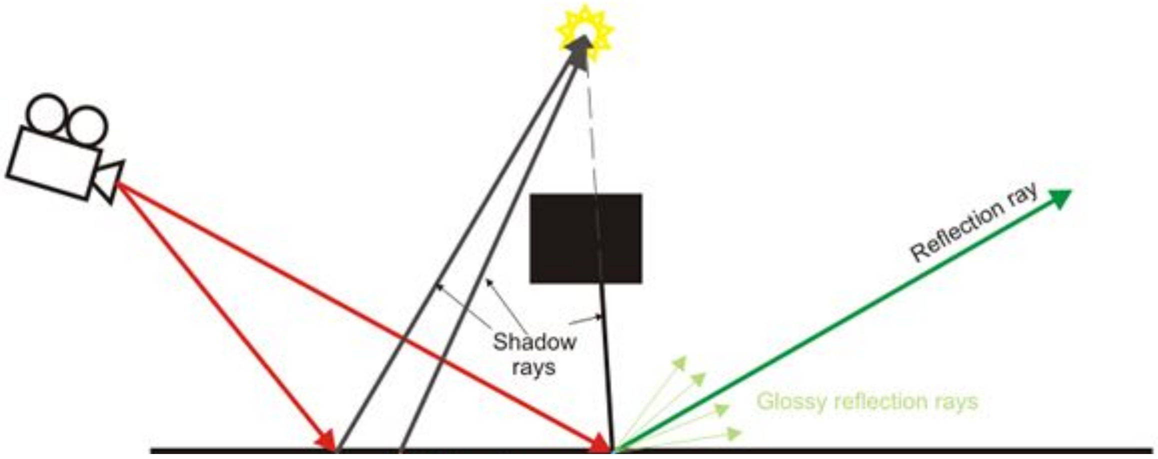
Geometry Query Acceleration

Billions of these queries are performed in a given DAGMC simulation.

A linear search over millions of triangles for every geometry query is untenable.

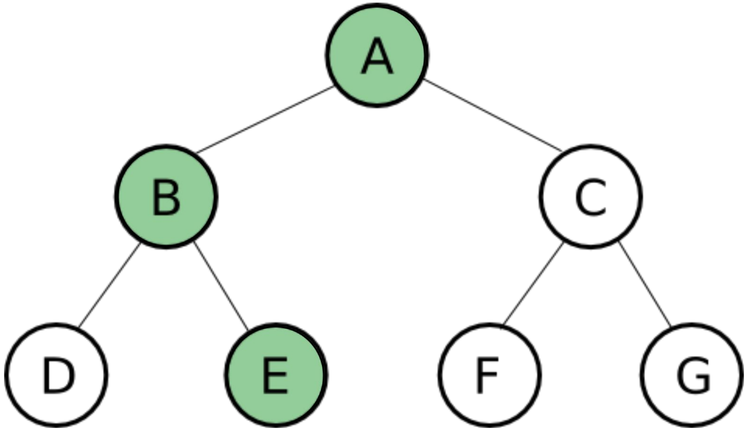
Ray Tracing

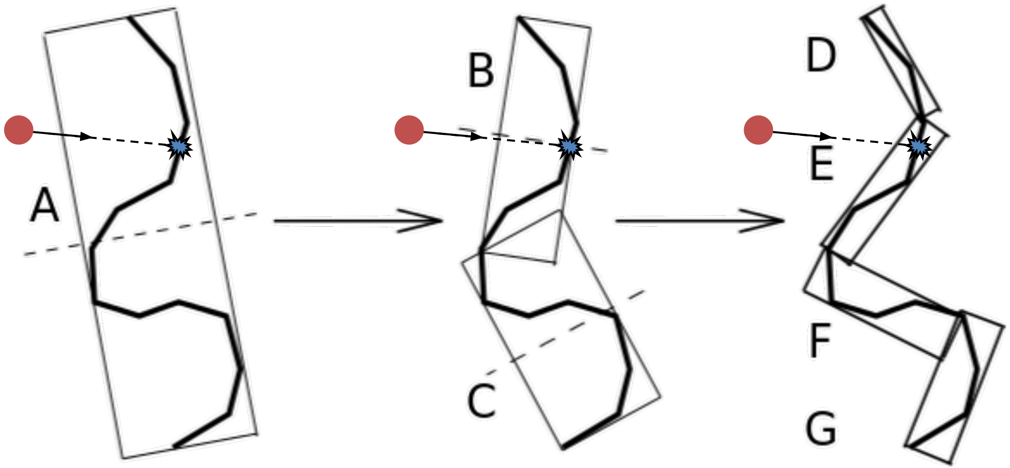
DAGMC currently employs the Mesh Oriented datABase, MOAB[3], to accelerate particle tracking through the triangle mesh using techniques from the field of ray tracing.



[7]

Bounding Volume Hierarchy (BVH)





[9]

Query Complexity:
$O(N) \rightarrow O(\log_2 N)$
N - number of triangles

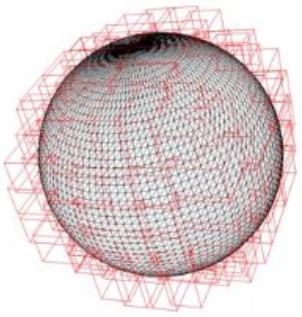
Aligned vs. Oriented Boxes

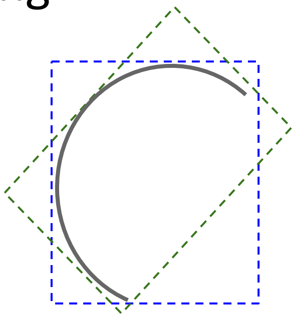
Axis-Aligned Bounding Box (AABB)

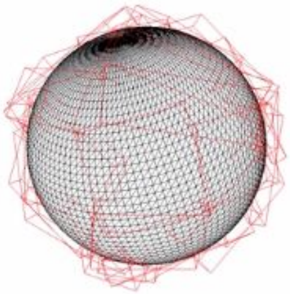
- easy to implement & debug
- less expensive intersection checks
- more overlaps, loosely fitting

Oriented Bounding Box (OBB)

- more difficult to implement
- more expensive intersection checks
- tightly fits primitives

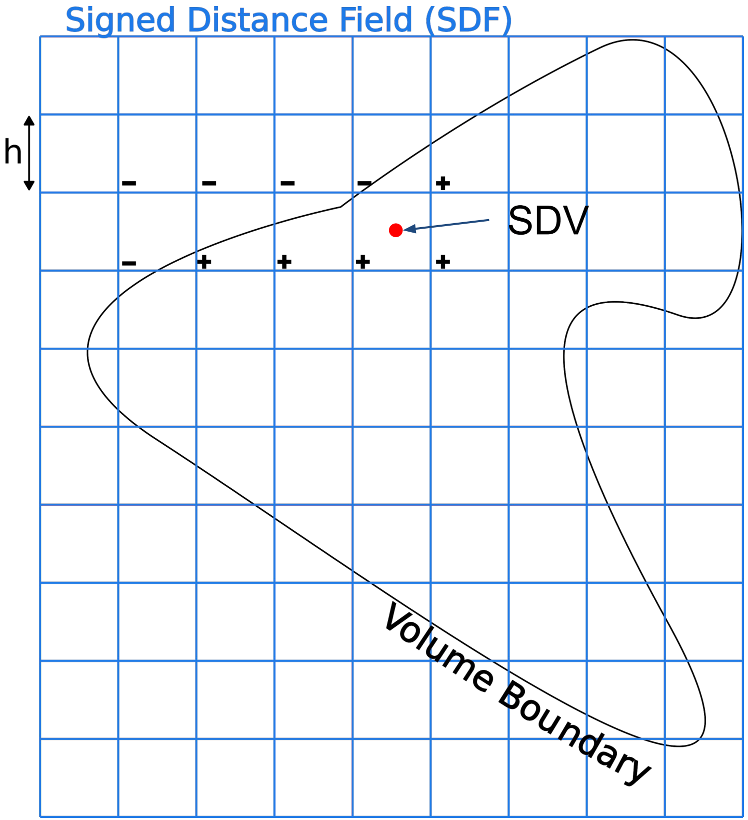






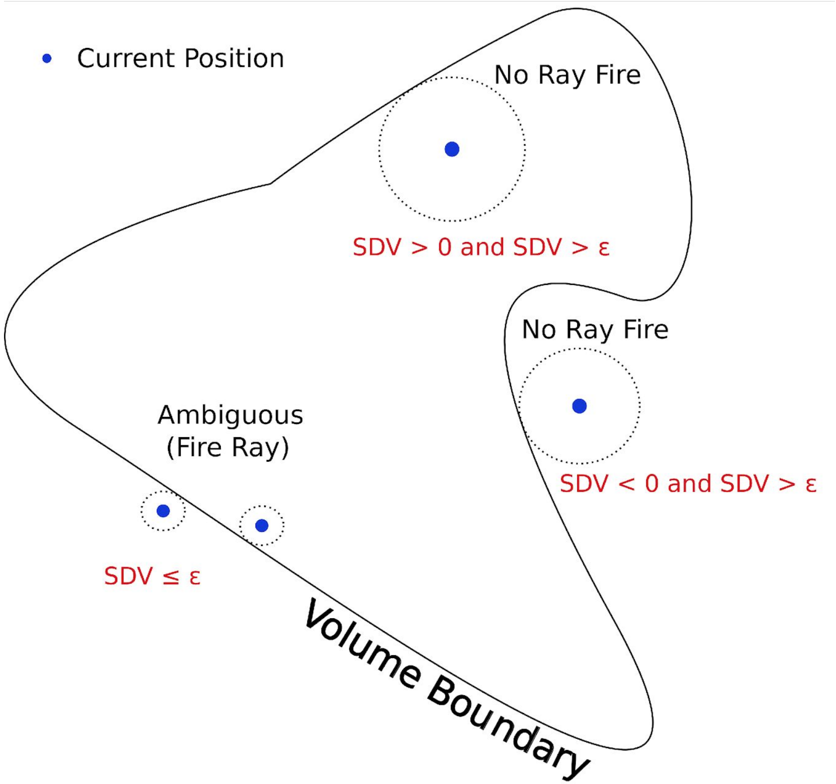
Signed Distance Field Preconditioning

Implicit Particle Tracking

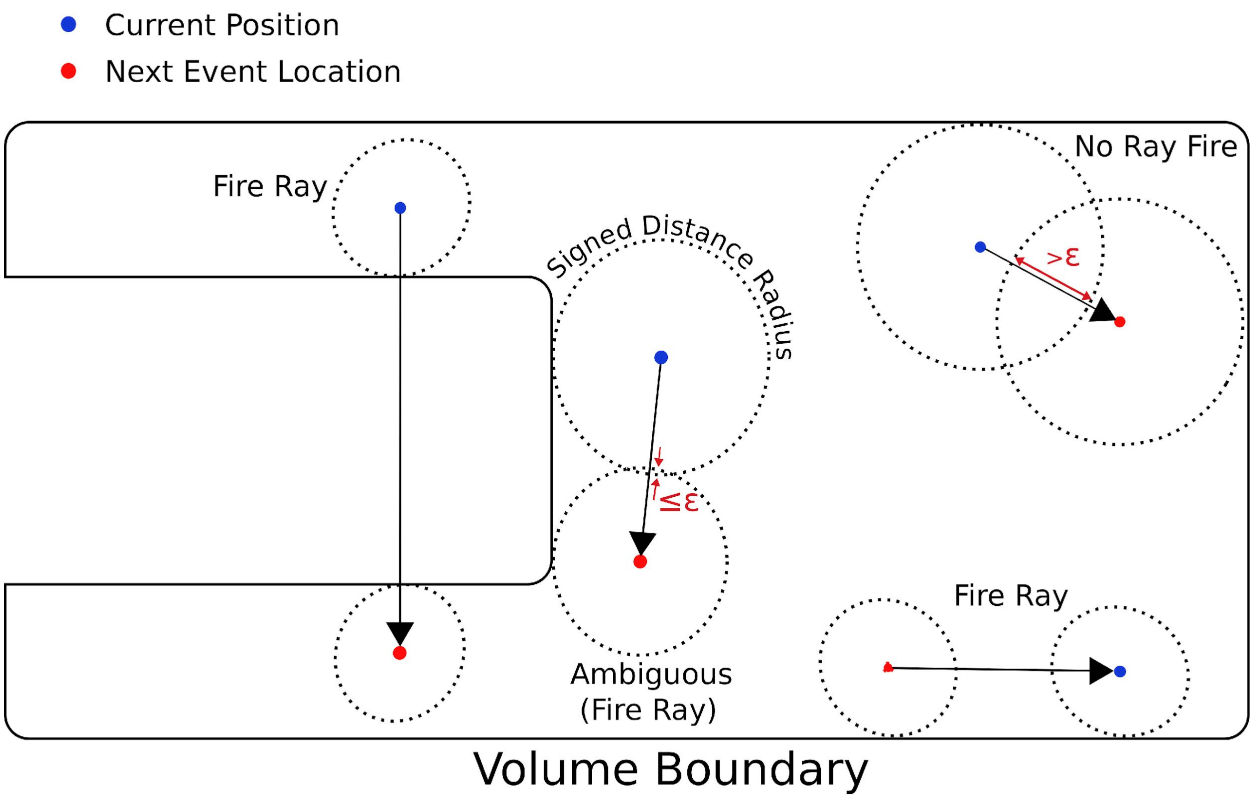


SDV - Signed Distance Value

Point Containment



Next Surface



Philosophy

- *From*: what surface would we hit next and how far is it?

  always $O(\log_2 N)$

- *To*: Could the particle hit a surface before the next physics event? If so, how which one and far is it?

  $O(1)$, then $O(\log_2 N)$

Preconditioner Utilization

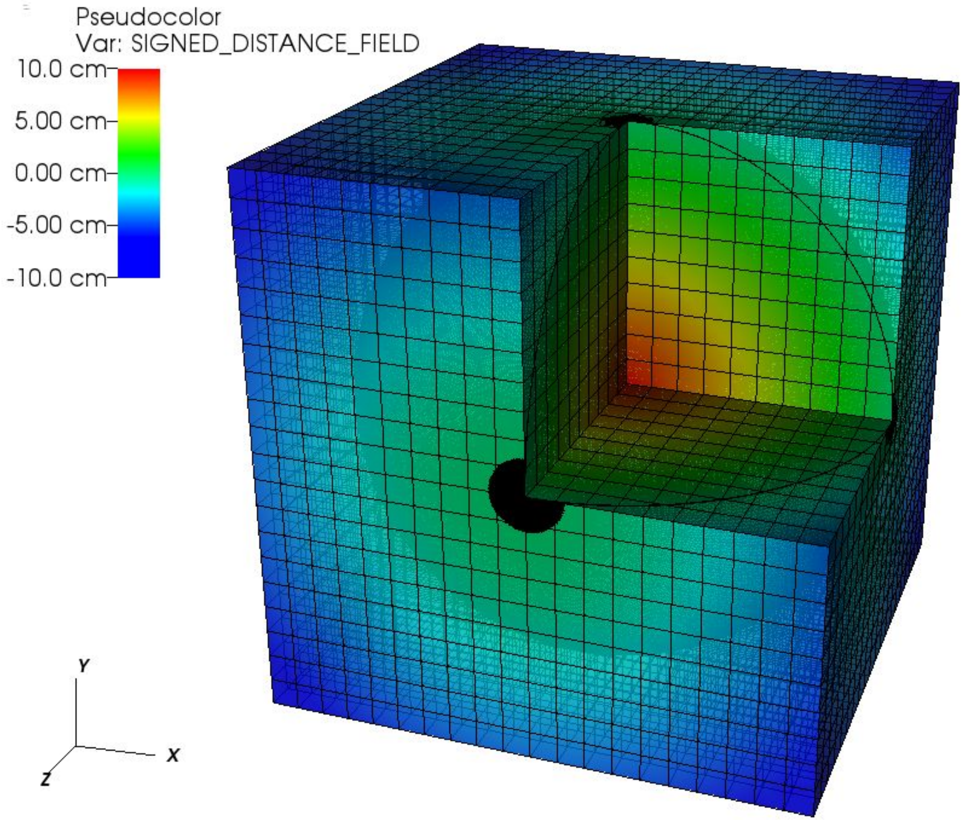
The utilization, U, of the data structure will vary depending on the problem scenario.

$$U = \frac{\text{Rays Avoided using SDF}}{\text{Ray Queries Performed}}$$

Goal: to create a predictive model for SDF utilization within a volume for any single physics event, based on some problem-specific parameters:

- mean free path
- characteristic volume size
- SDF mesh step size

SDF Visualization



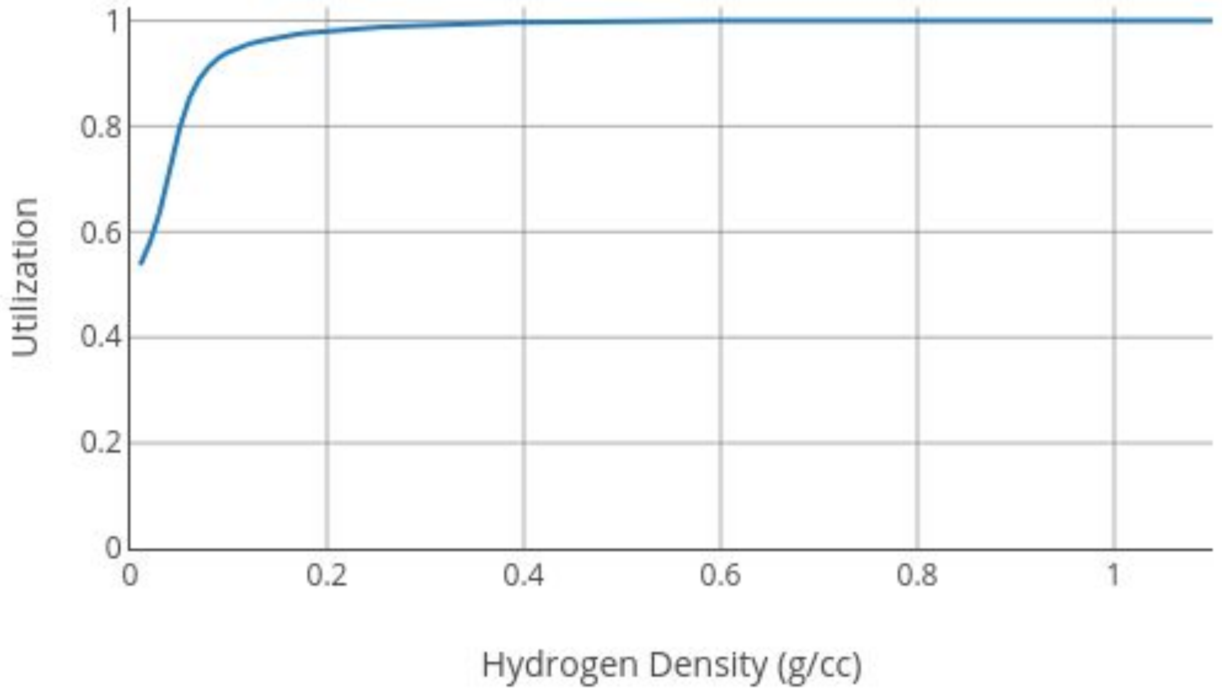
Testing

Model:

- 10 cm radius dense hydrogen-filled sphere centered on the origin
- 5 MeV neutron point source at the origin
- histories in each run: 1M
- SDF mesh step size: h = 0.5 cm
- triangles in the model: ~1M

Implementation	Wall time (min)	Ratio
MCNP5	2.75	1
DAG-MCNP5	25.04	9.13
DAG-MCNP5 SDF	7.15	2.65

Preconditioner Utilization

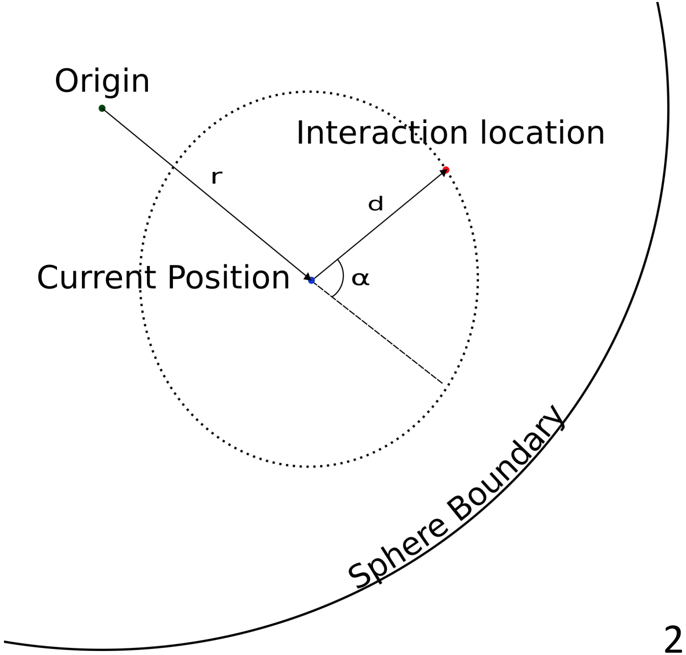


Utilization Modeling

Simple Monte Carlo tool used to develop a predictive model:

- Particles were sampled uniformly with position, r, in a spherical geometry and scattered once through an angle, α.

- The distance to the next physics event, d, was then sampled using two distributions governed by a pseudo-physical input parameter, λ:
  - Fixed Distance: $$p(d) = \frac{\delta(d-\lambda)}{d^2}$$
  - Sampled Distance: $$p(d) = \frac{e^{-\frac{d}{\lambda}}}{d^2}$$



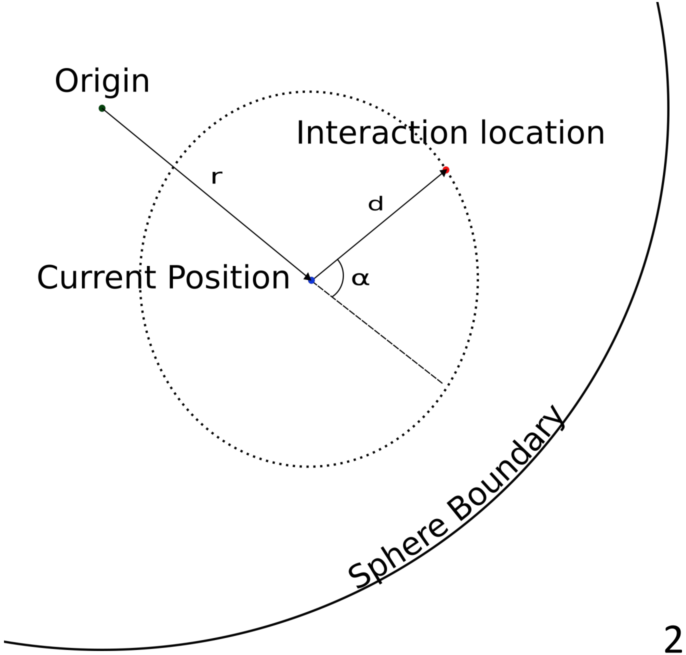
Utilization Modeling

Analytic model:

$$\int_{V_{sphere}} \int_{V_{track}} p_p(r) p_n(d) dV_{sphere} dV_{track}$$

$p_p(r)$ - probability of particle's source location, r (uniform)

$p_n(d)$ - probability of particle's next event location for distance, d, and angle, α



Modeling Parameters

Model:

h - mesh step size

ϵ - interpolation error

λ - mean free path

R - sphere radius

Plots:

R_{av} - average chord length

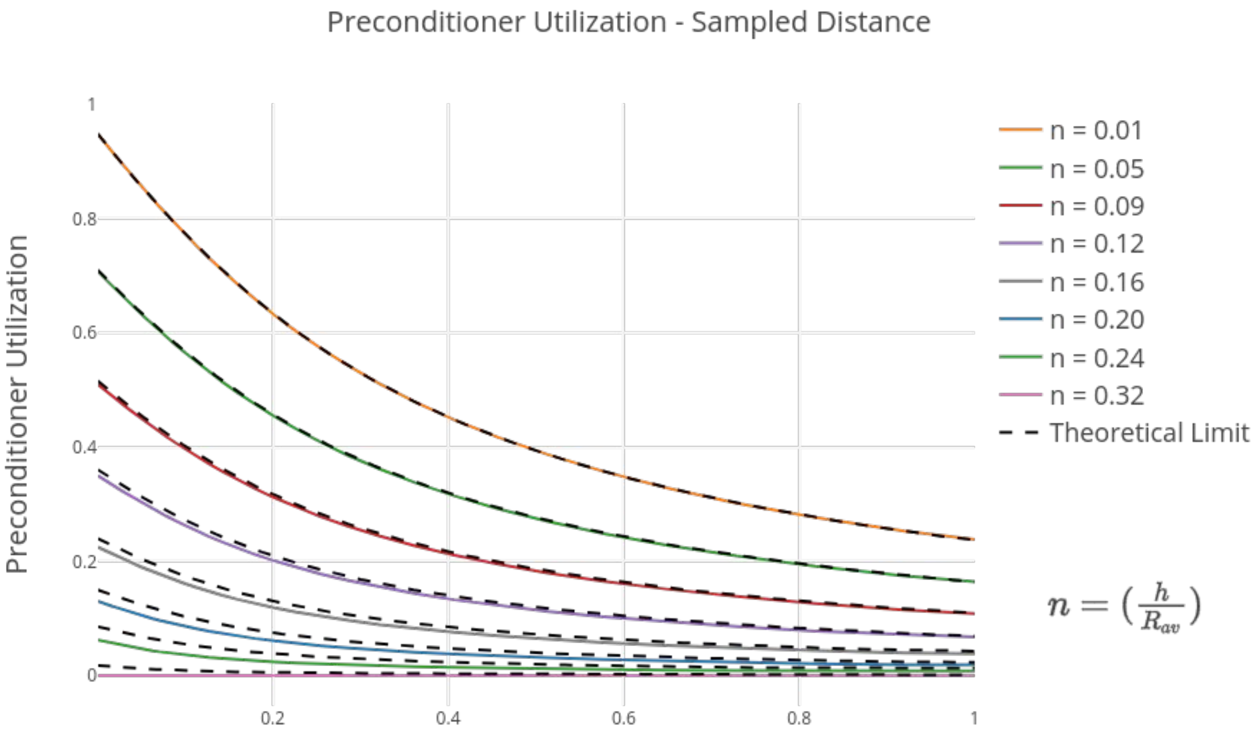
$$R_{av} = \frac{4V}{S} = \frac{4}{3}R$$

$c = \lambda / R_{av}$

$n = h / R_{av}$

$$U = \frac{(R-\epsilon)\left(\frac{1}{2}\lambda(R-2\lambda-\epsilon)e^{\frac{-R+\epsilon}{\lambda}} + \lambda^2 - \frac{3}{2}\lambda(R-\epsilon) + (R-\epsilon)^2\right)}{R^3}$$

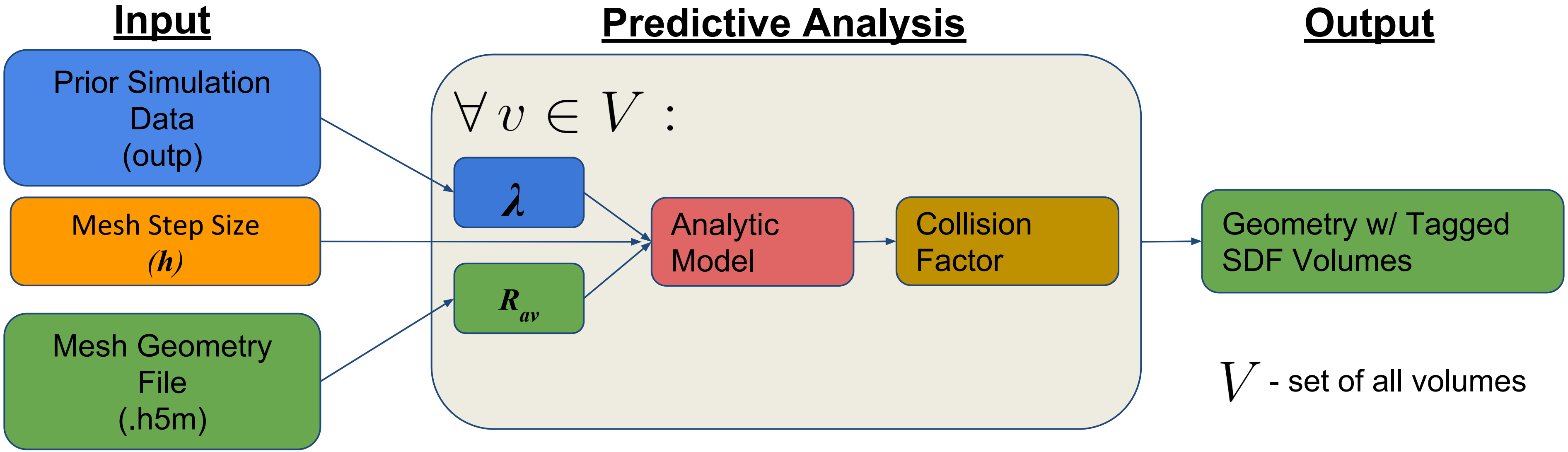
Utilization Modeling



$$n = \left(\frac{h}{R_{av}}\right)$$

$$c = \left(\frac{\lambda}{R_{av}}\right)$$

Model Application

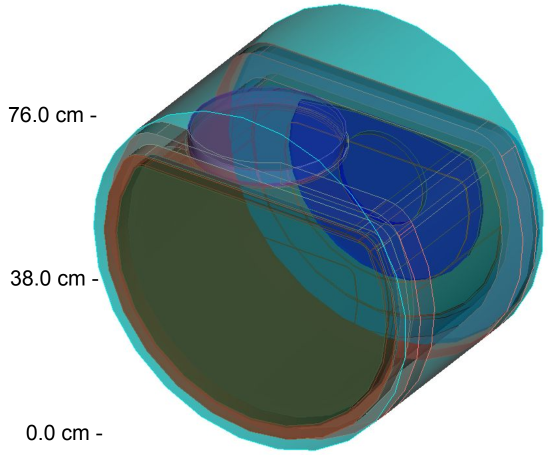


V - set of all volumes

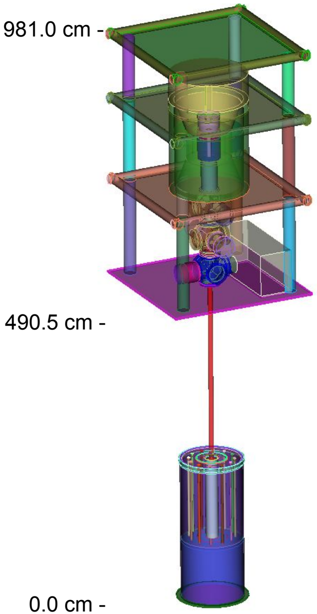
Collision Factor:

$$\frac{\text{Physics Events in Volume}, v}{\text{Total Physics Events in Simulation}}$$

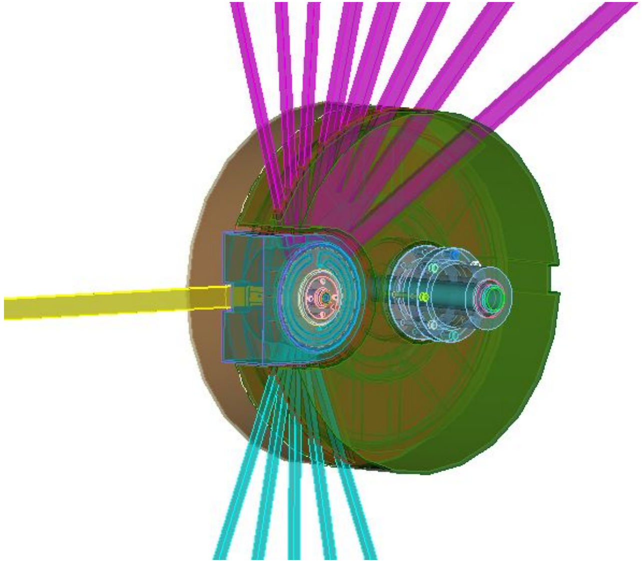
New Production Models



neutron Time of Flight
(nTOF)



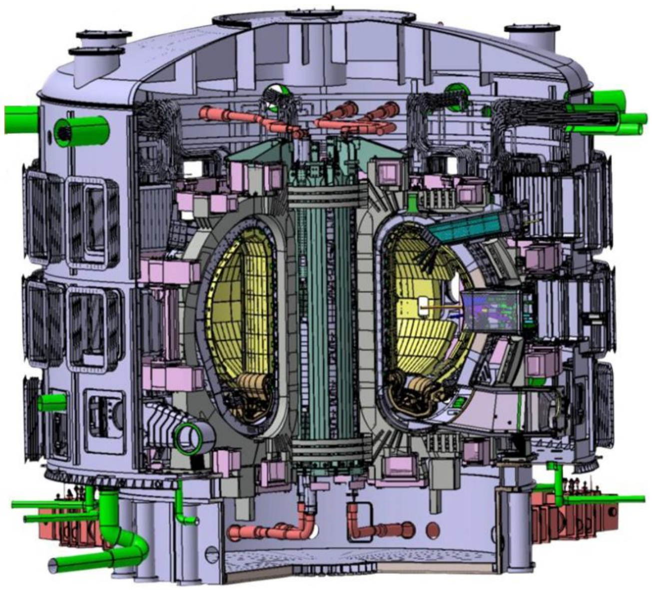
SHINE

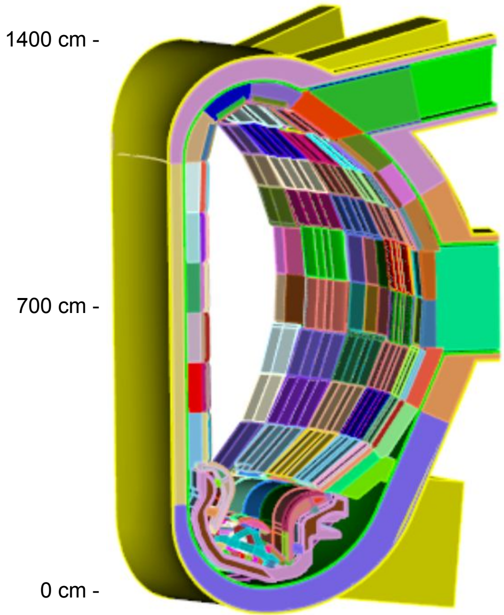


70 cm x 70cm x 200 cm

Spallation Neutron Source
(SNS)

ITER



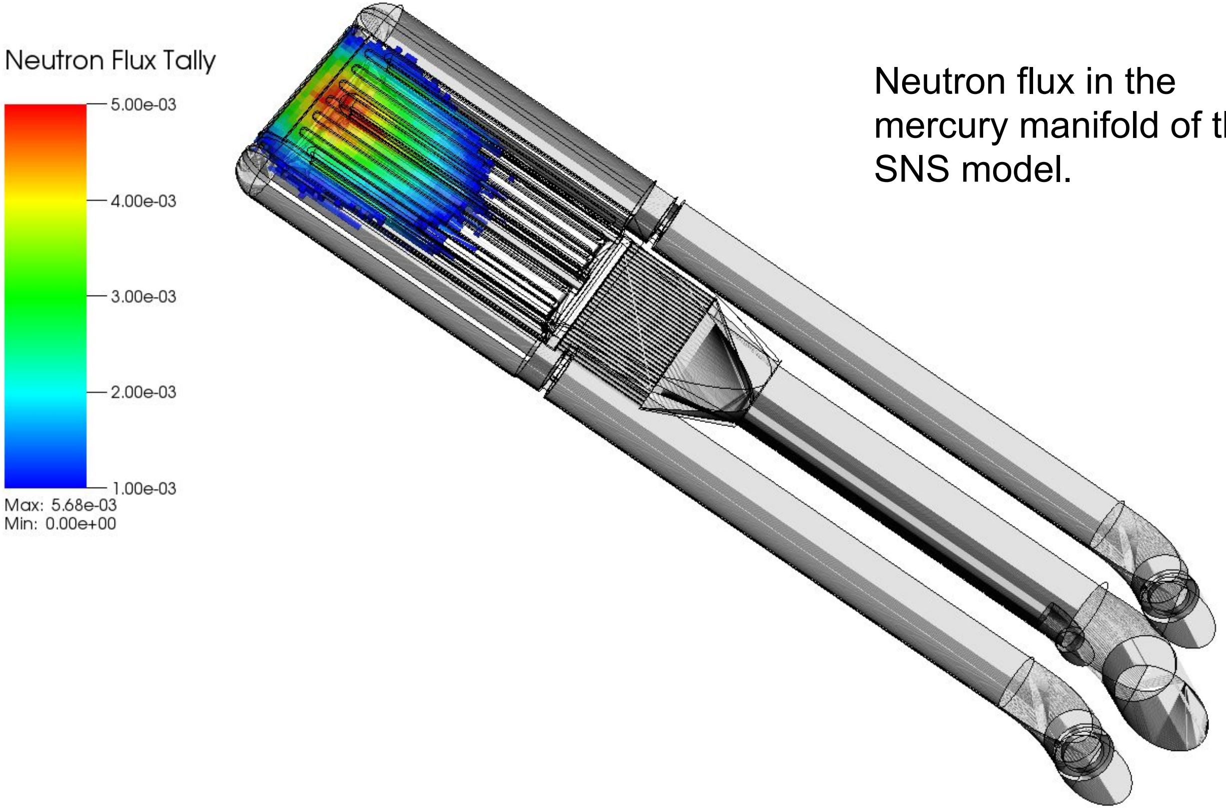


Production Results

Model	Predicted Utilization (%)	Actual Utilization (%)	Timing Ratio
ITER	37.1	40.1	1.16
nTOF	62.9	66.2	1.41
SHINE	60.1	48.2	1.19
SHINE*	60.1	56.5	1.56
SNS	31.0	2.1	~1.00

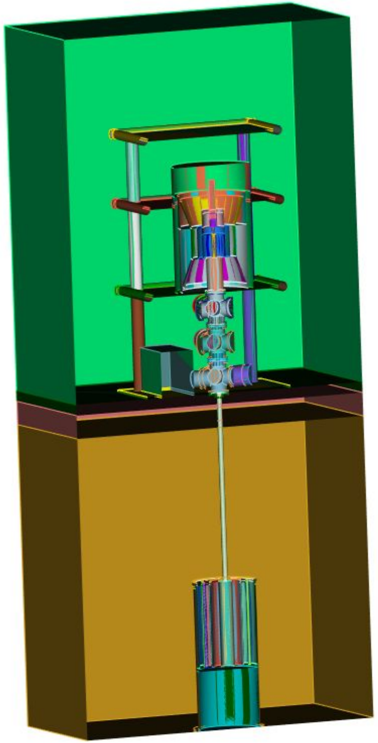
* - SHINE's cross-section data modifications

Predictive Model Limitations



Neutron flux in the mercury manifold of the SNS model.

SHINE Application



SHINE Model

Resulted in a 36% reduction in compute time with identical tally results for 500k particles

$$\frac{\text{DAGMC w/ SDF}}{\text{DAGMC}} = \frac{49.28min}{76.14min} = 0.64$$

Extended to a 100M history run, this results in **~91 hrs** of saved computing time.

Summary

- Moderate improvements in simulation run times were observed with no change in the numerical results from DAGMC.
- The predicted utilization from the analytic model matched the measure utilization for most simulations.
- Various problem factors and model limitations diminished the effect of SDF preconditioning in production simulations.

Future Work

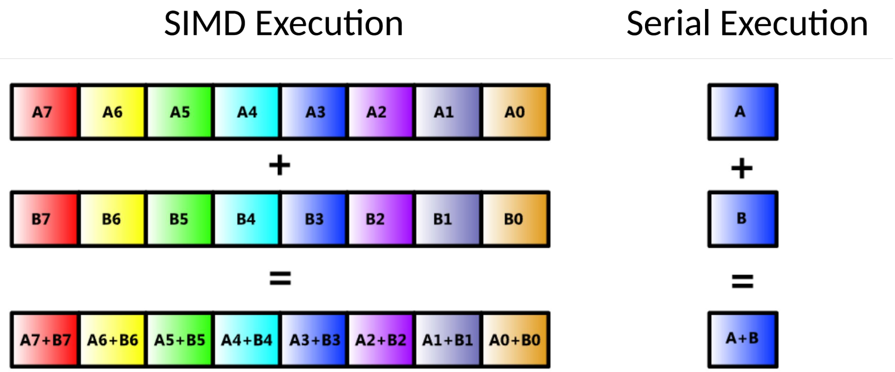
Future Work:

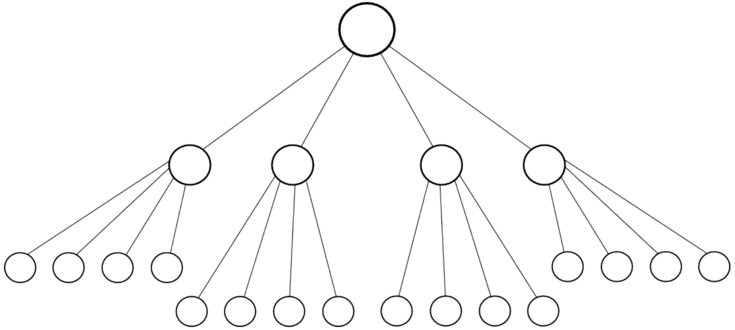
- Implementation with other physics codes, analytic model improvement, and alternative mesh storage structures
- Application of predictive model in other methods such as Woodcock delta tracking[12]

Mixed Precision Ray Tracing

Intel’s Embree

- Embree[6] uses single-precision values for triangles and BVH
- use of single instruction multiple data (SIMD) commands
- SIMD-oriented BVH hierarchy design
- Axis-Aligned Bounding Boxes

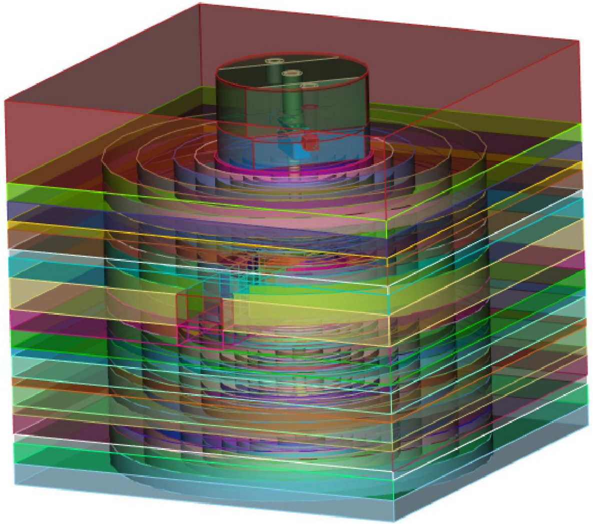




EmDAG Production Model Test

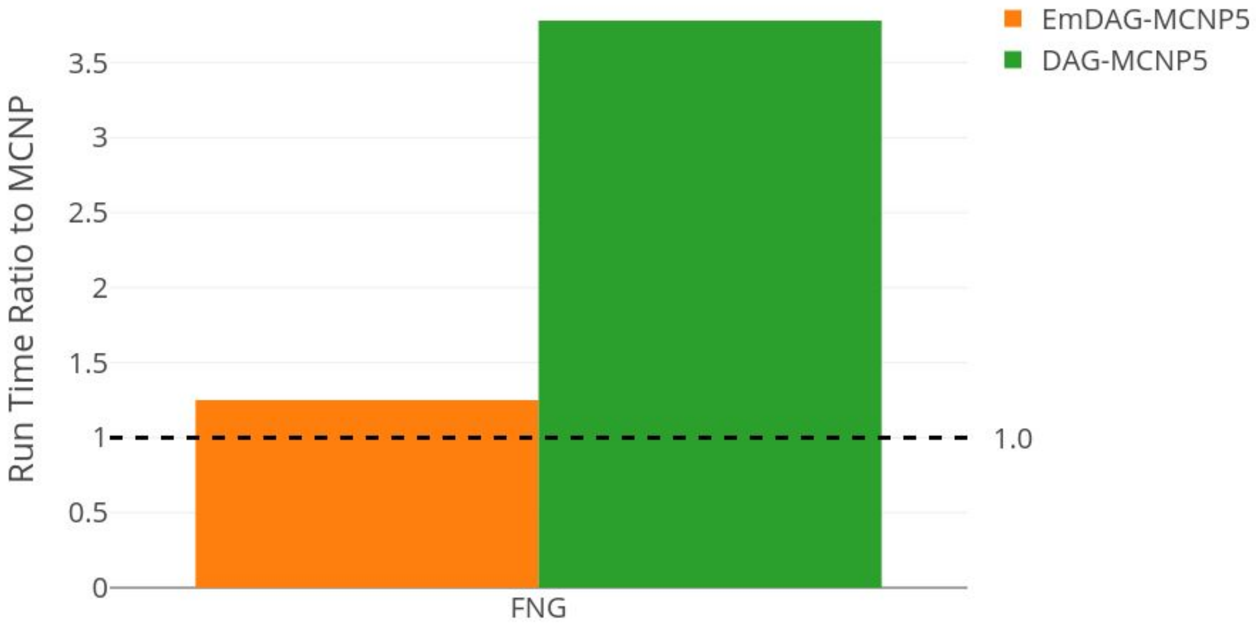
FNG Test:

- volumetric source
- 100M 14.1 MeV source particles



EmDAG Production Model Test

FNG Test
(volumetric source)



Native & DAGMC: zero lost particles
EmDAG: 240 lost particles in 100M

Motivation

EmDAG Conclusions:

- Successful demonstration of data parallelism for improved simulation times
- Single to double precision conversions can decouple numerical and logical positioning

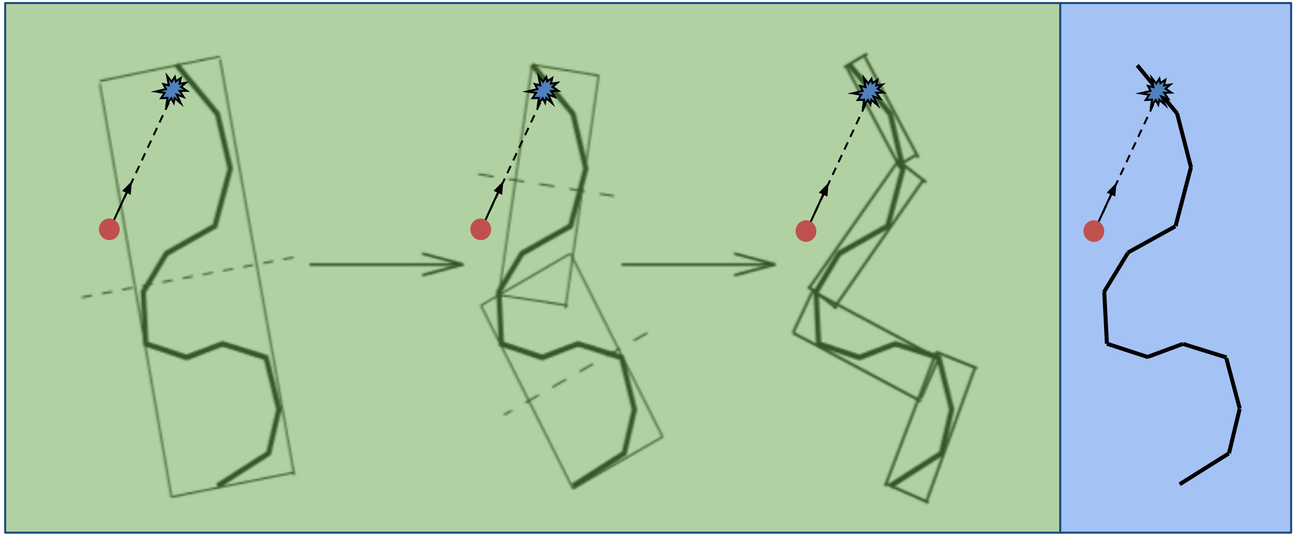
Moving forward:

- Embree-based traversal of BVH data structures
- Return of double precision values using MOAB's Plücker intersection[11] method for consistency

Proposed Design

True/False

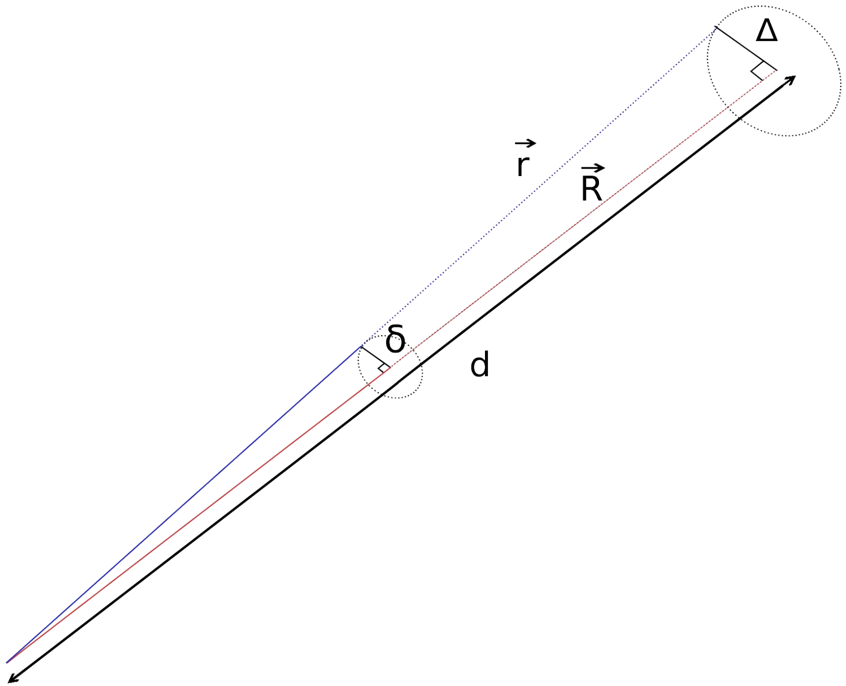
Return Value



single
precision

double
precision

Box Extension Parameter



$\vec{R}$ – double precision ray

$\vec{r}$ – single precision ray

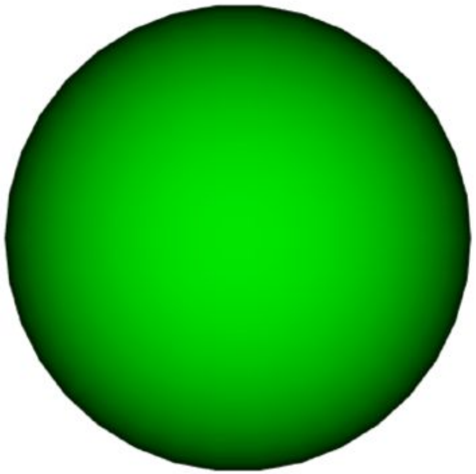
d_{max} – longest chord

$\delta = 10^{-7}$ [10]

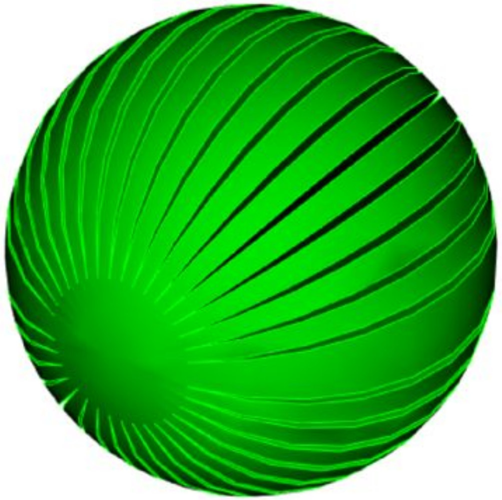
$$\Delta_{max} = \sqrt{3} \times 10^{-7} d_{max}$$

Note: the value of δ may vary depending on CPU.

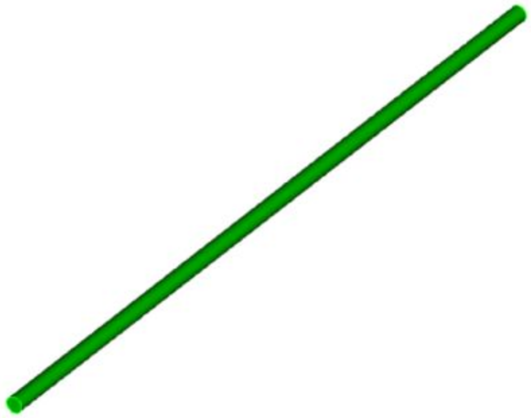
Ray Fire Test Models



Sphere

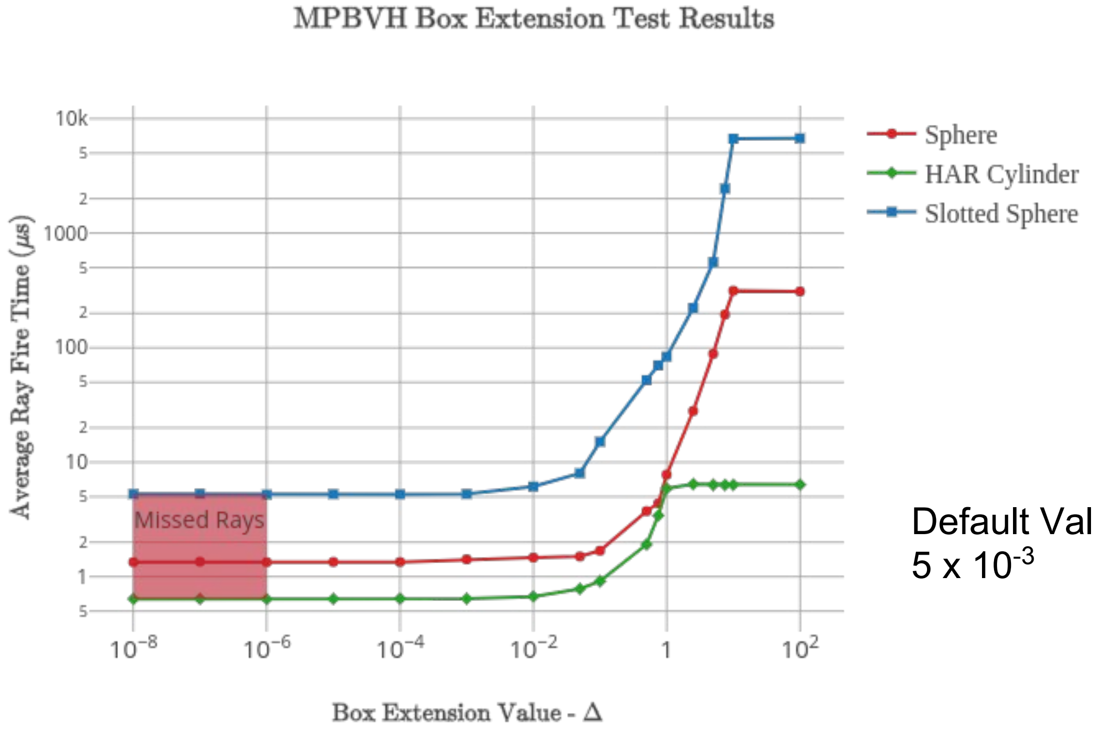


Slotted Sphere



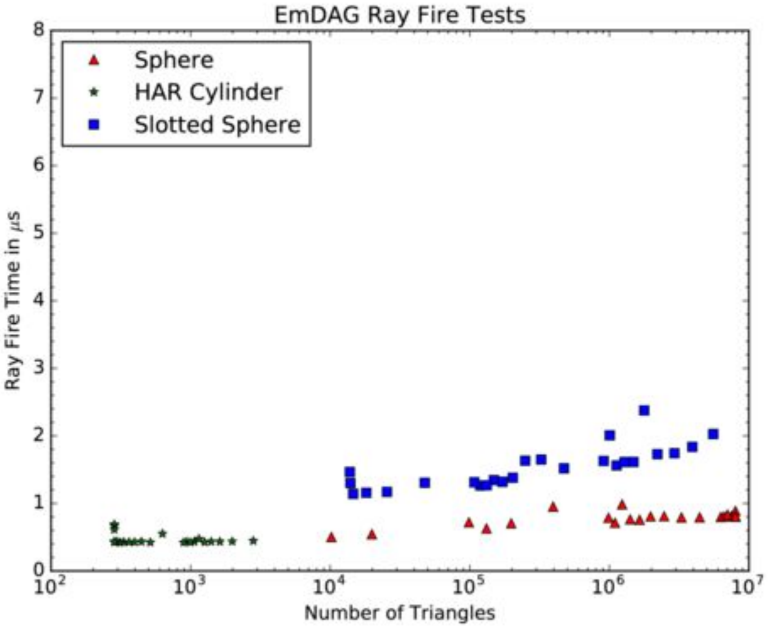
High Aspect Ratio Cylinder

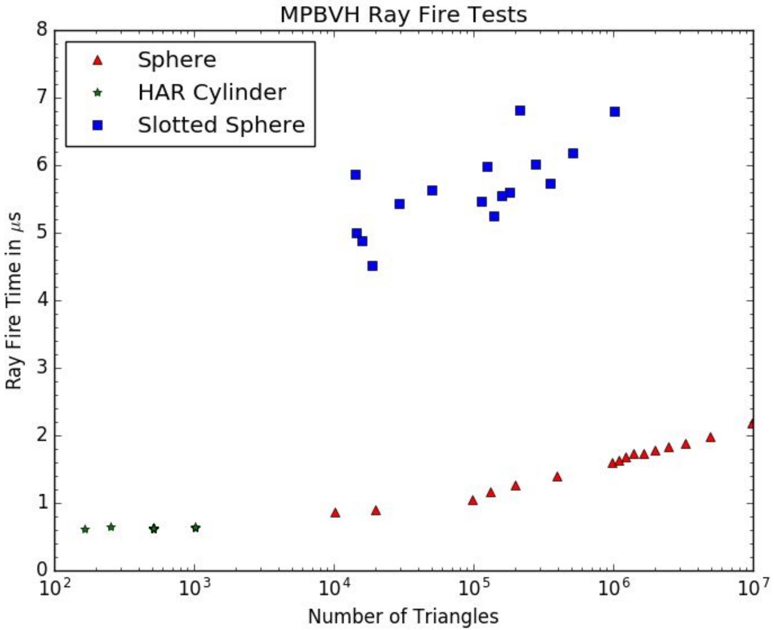
Box Extension Study



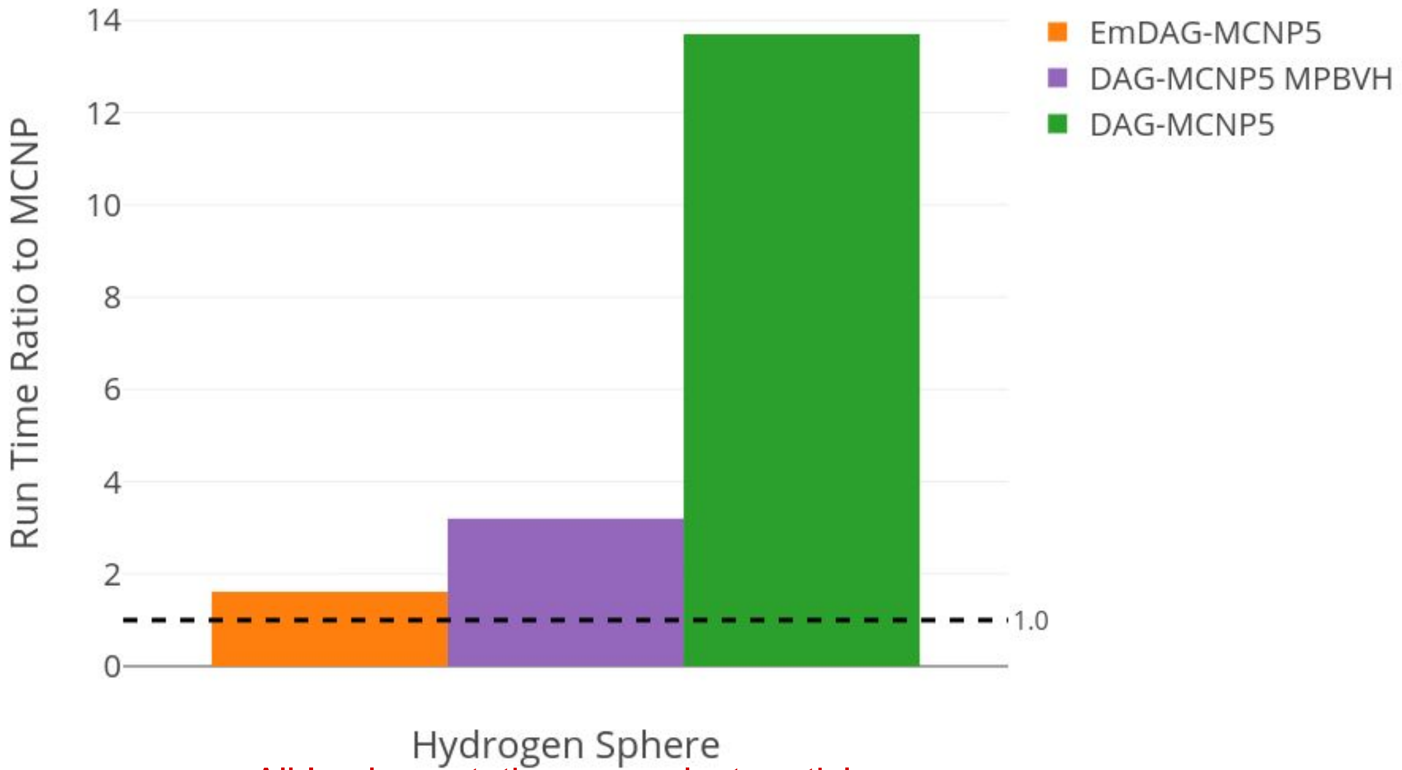
Default Value:
5×10^{-3}

Ray Fire Performance



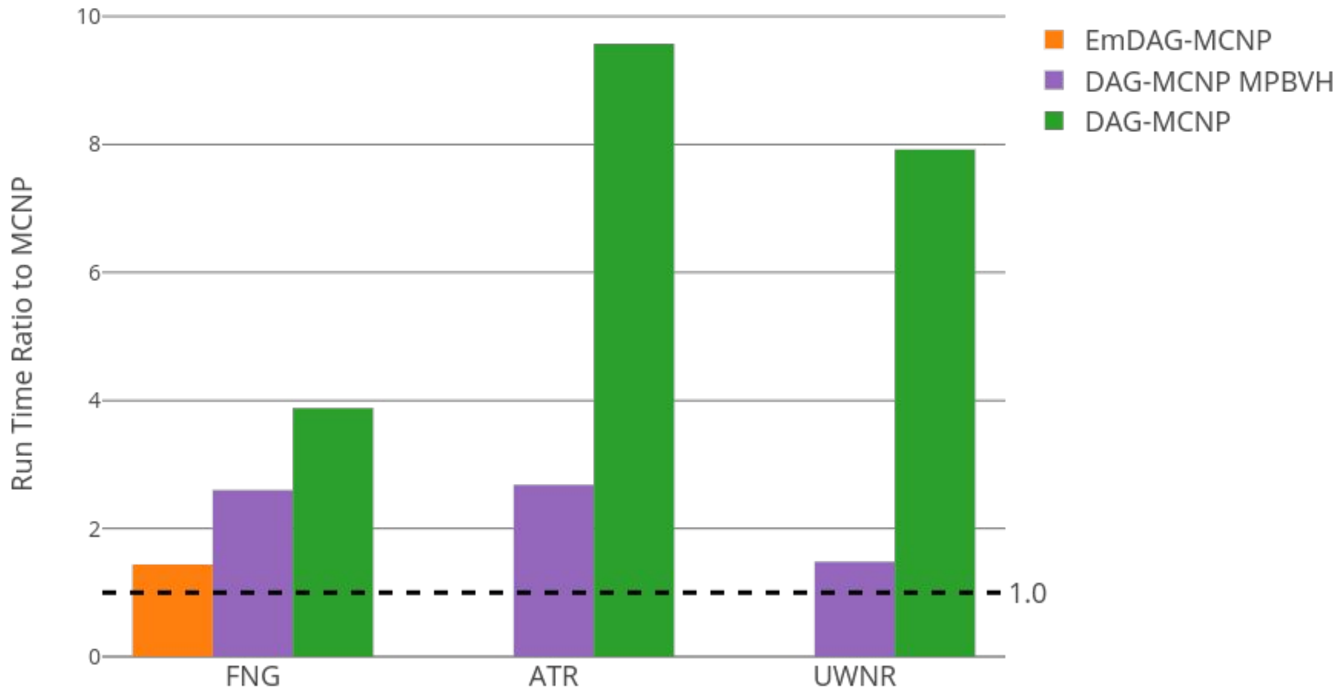


Initial Results



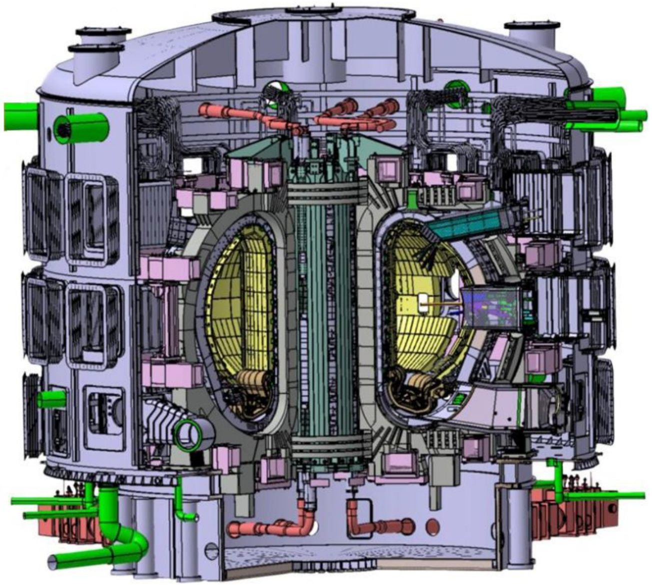
All Implementations: zero lost particles

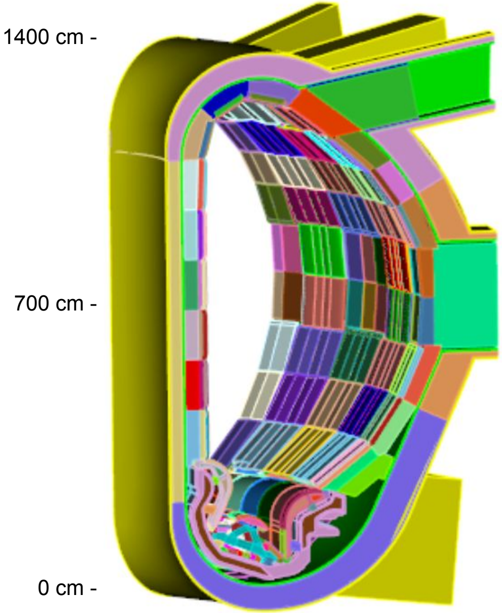
Production Results



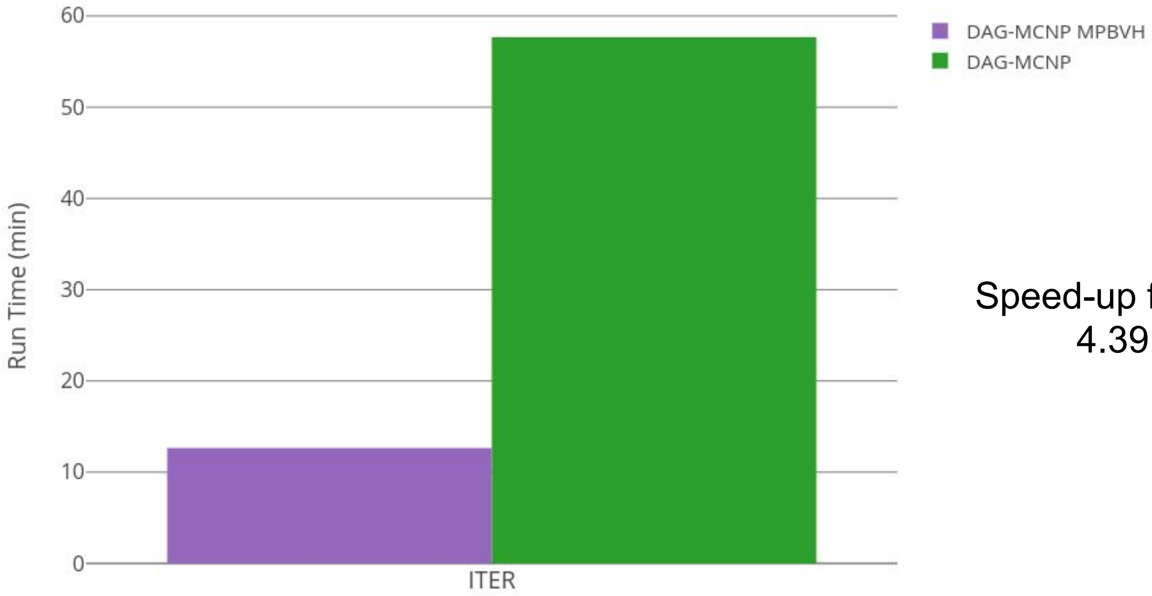
All MPBVH Implementations: zero lost particles, same numerical answer

ITER





ITER Performance



Speed-up factor:
4.39

10M Particles

zero lost particles, same numerical answer

Summary

Use of reduced precision bounding volumes and vectorized traversal resulted in improved performance anywhere from 2-5x faster than an unmodified version of DAGMC while producing the same numerical result in all test cases.

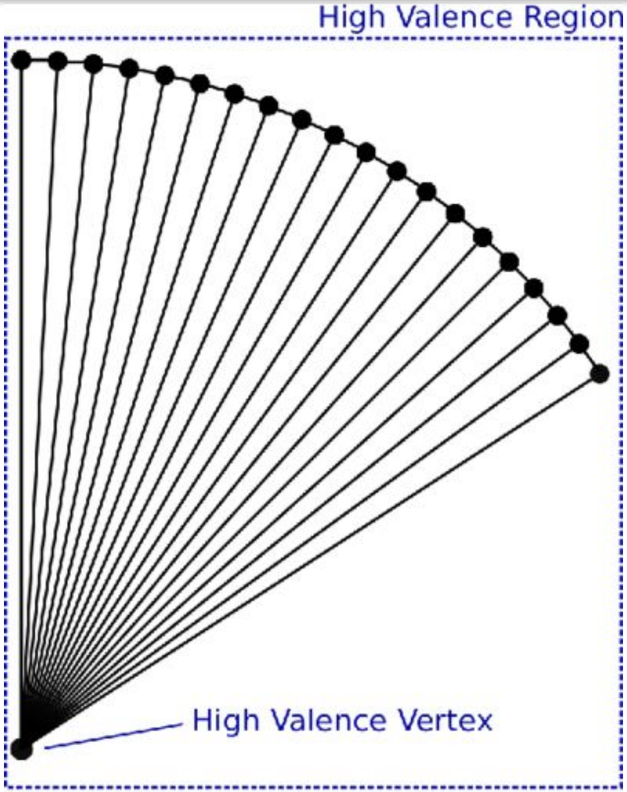
Future Work

Future work:

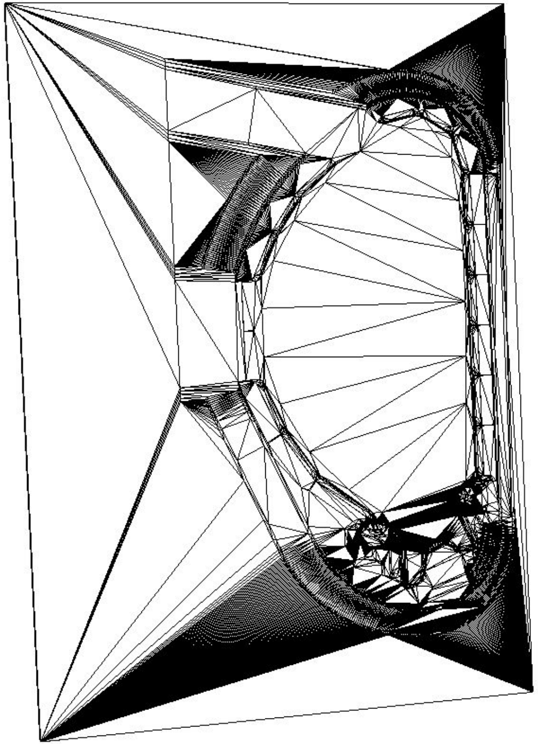
- Additional work for management of the box extension value on a per-volume basis could be useful in larger models.
- Extensions of this work to more exotic CPU architectures (Xeon PHI) would theoretically provide further performance improvement.
- Other nascent methods for managing reduced precision intersections in large models could be explored as well.[8]

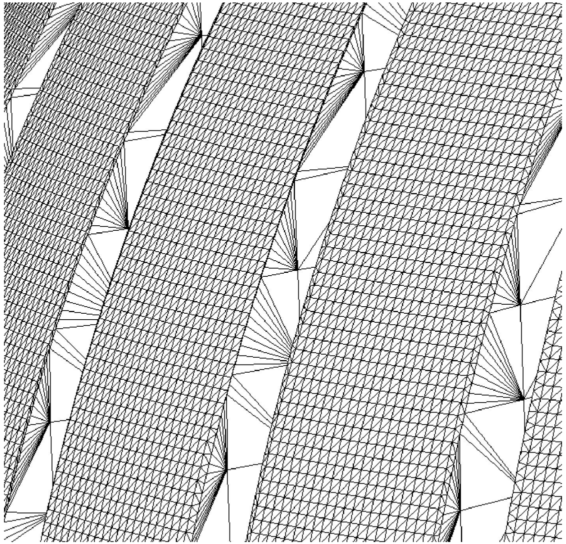
Mesh Feature Adaptation

High Valence Regions

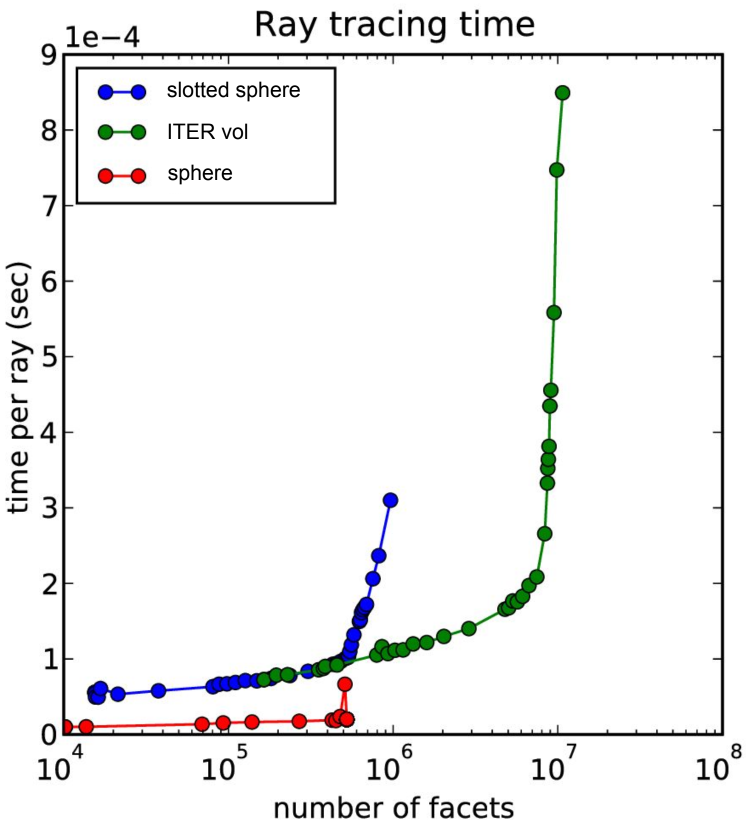


High Valence in Production



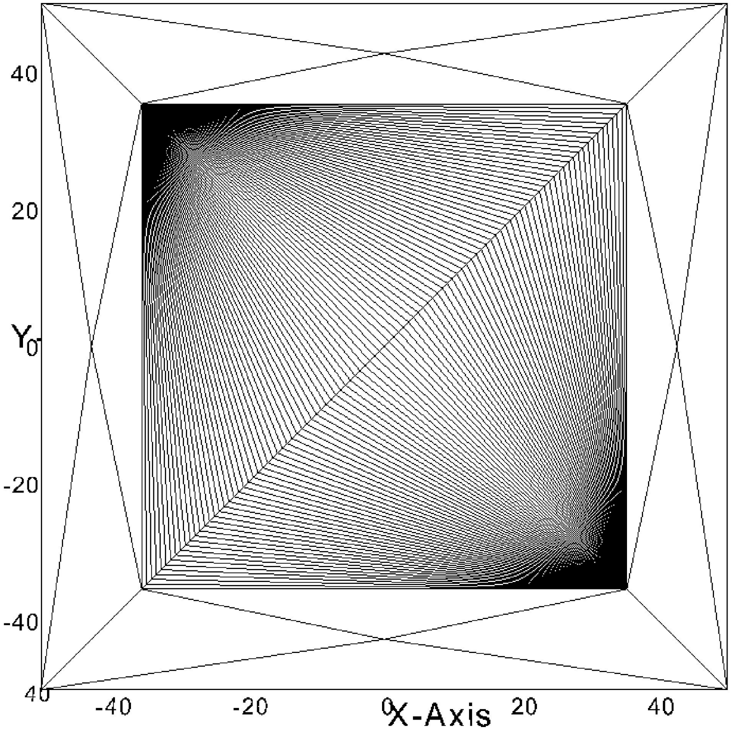


High Valence Performance

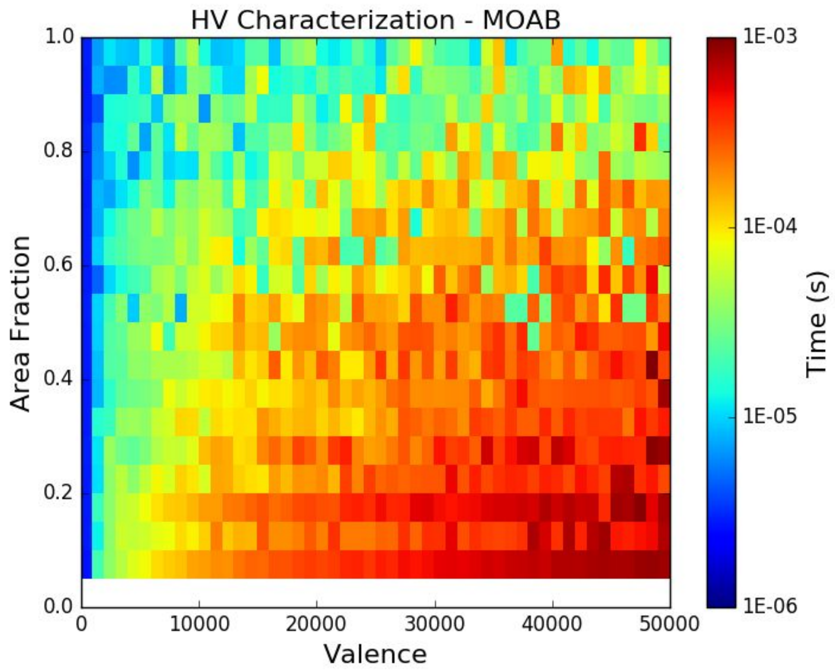


[11]

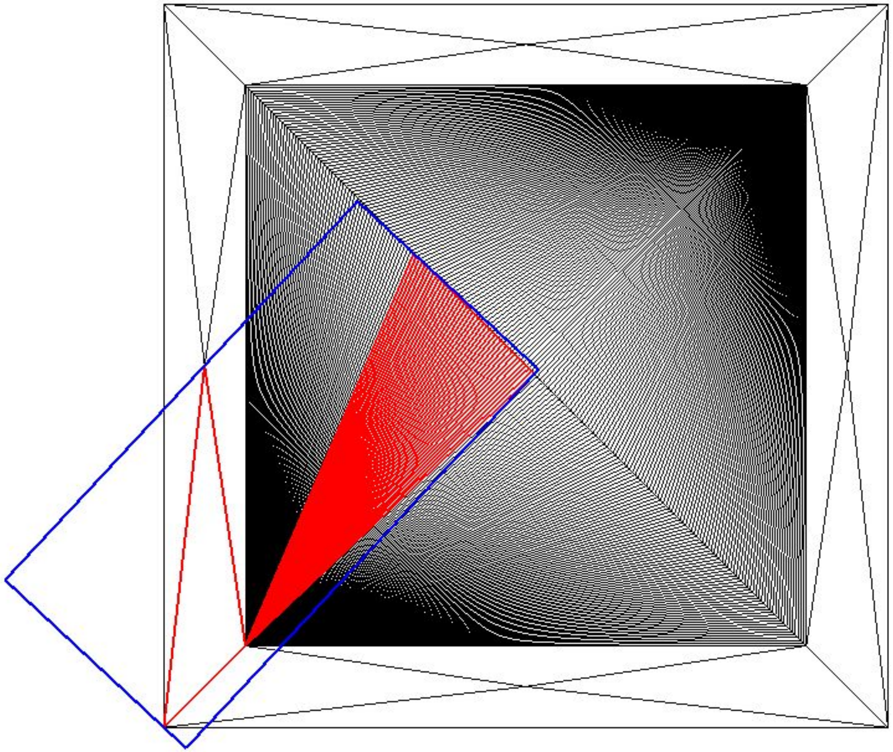
High Valence Test Model



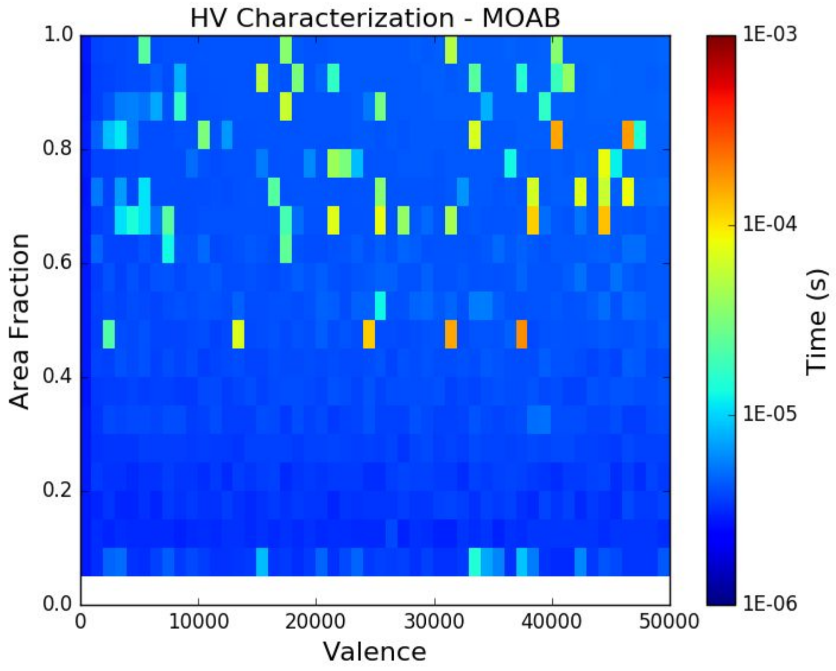
HV Study Timing Results: DAGMC



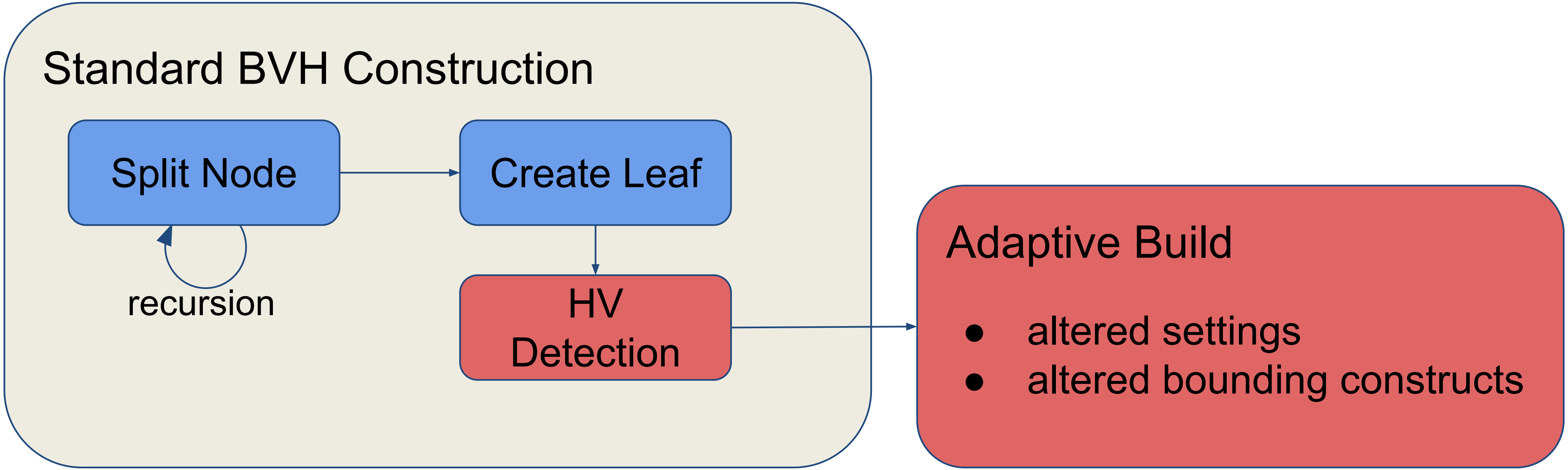
High Valence Region: MOAB Leaf Nodes



HV Study Timing Results: DAGMC



Adaptive Construction



High Valence Detection

High Valence Condition:

$$\frac{|A_{v,tri} \bigcap T|}{|T|} > \alpha \, \forall \, v \in V$$

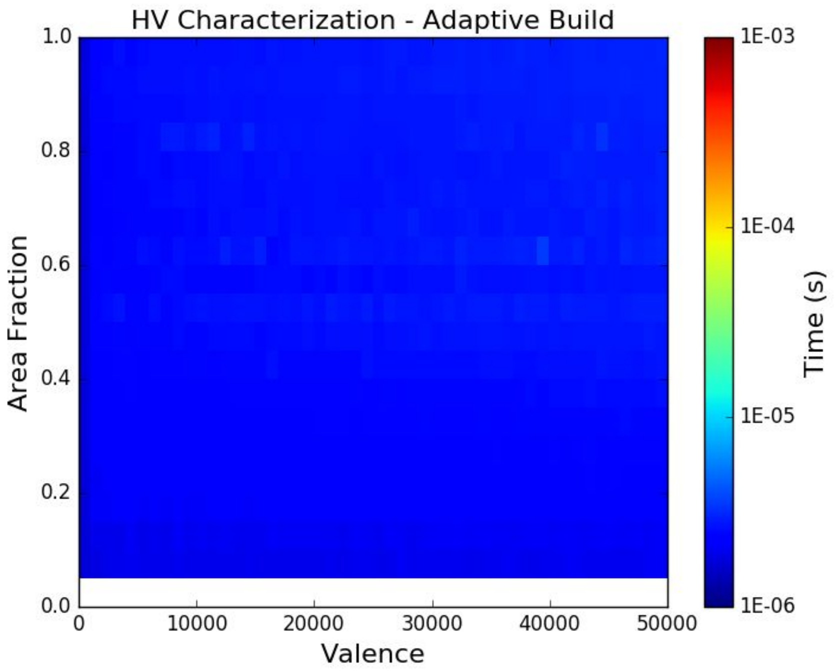
$A_{v,tri}$ - adjacent triangles to vertex, v

T - set of triangles in the bounding volume

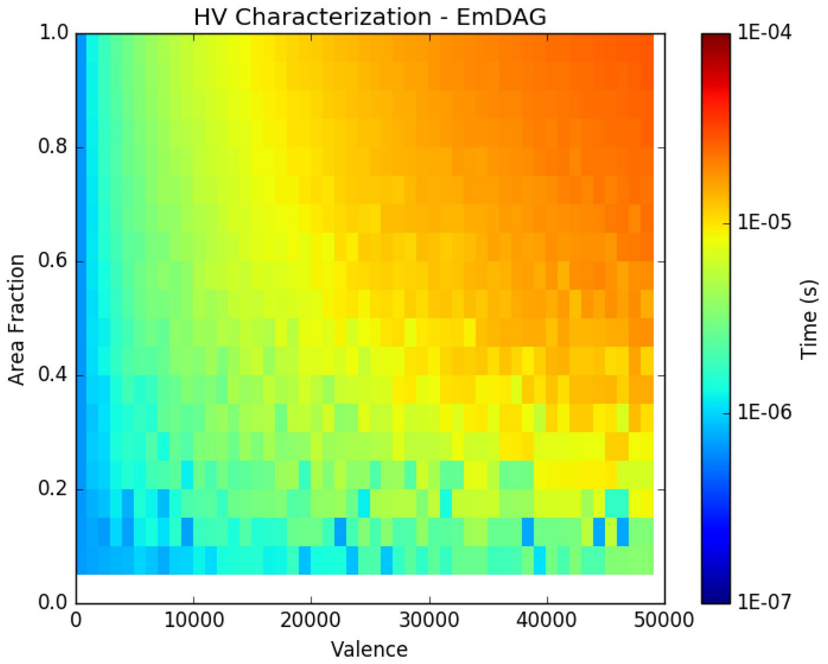
α - high valence parameter

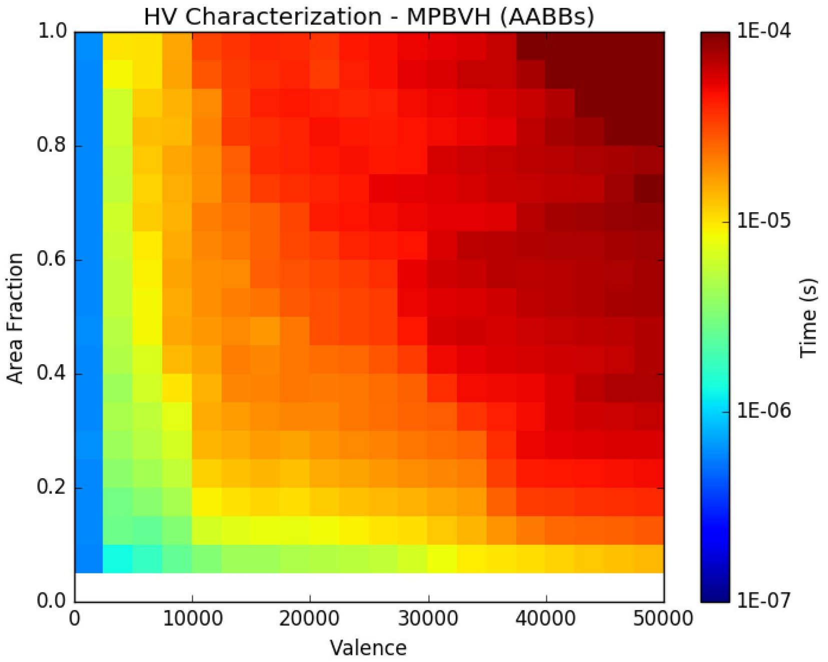
V - set of vertices in the bounding box

HV Study Timing Results: Adaptive Build

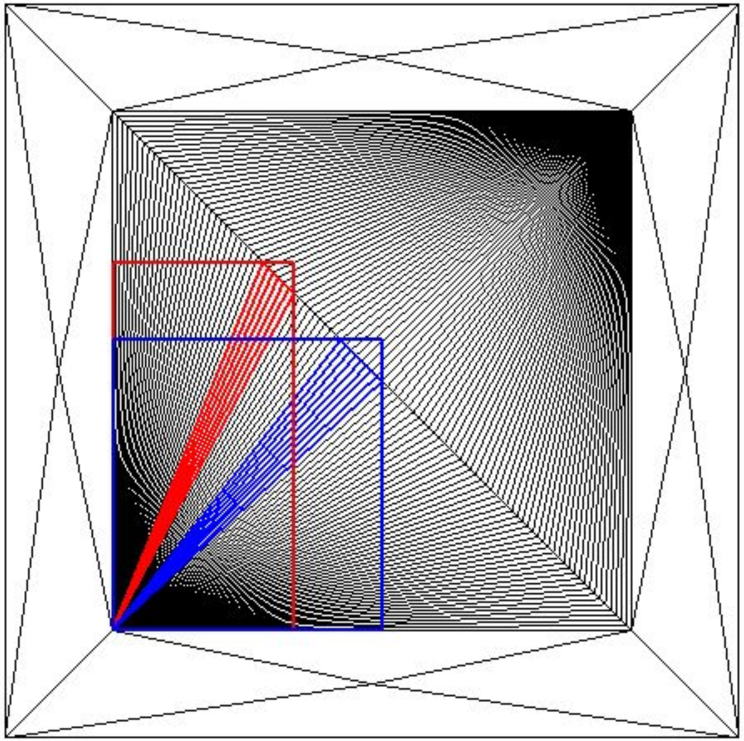


HV Study Timing Results: EmDAG & MPBVH

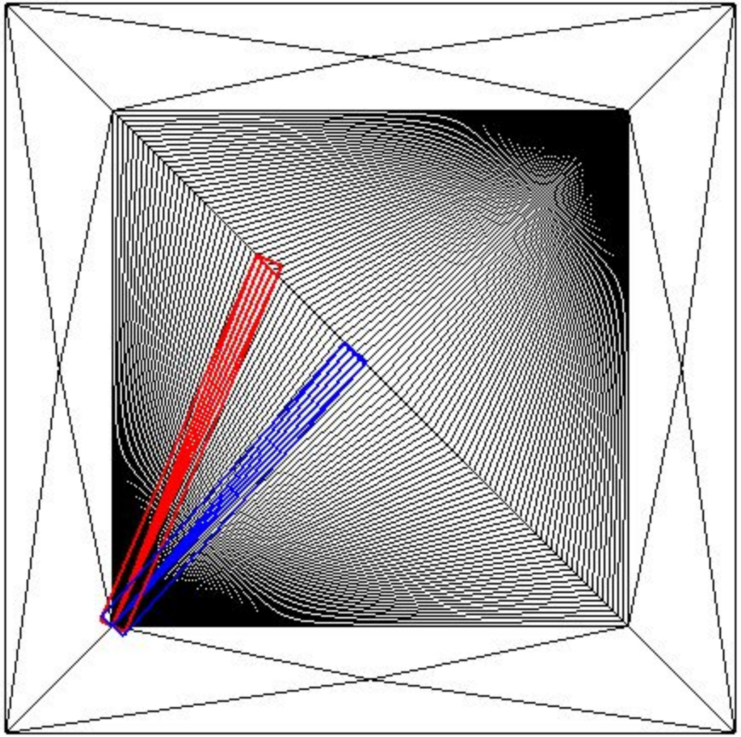




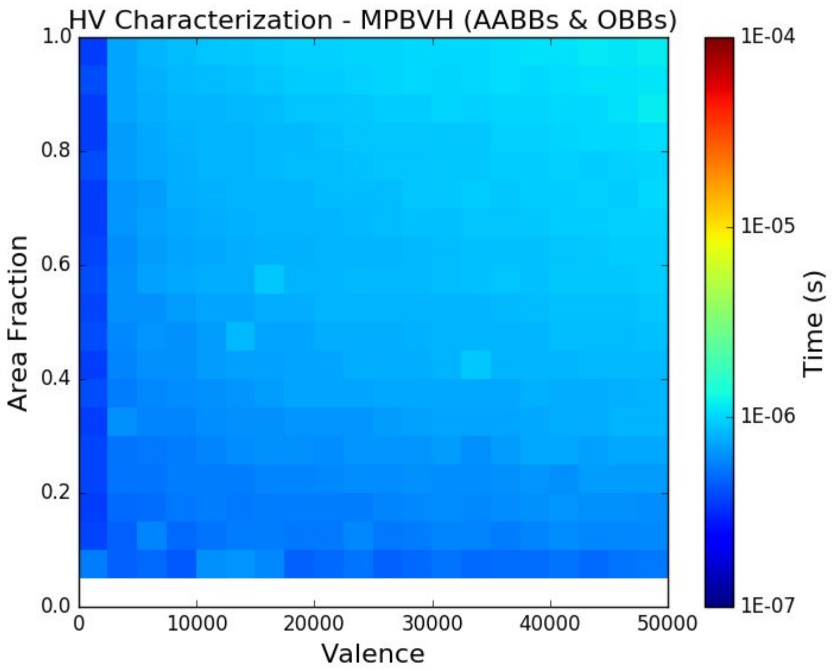
High Valence Region: MPBVH Leaf Nodes



OBBs in High Valence Regions



MPBVH Results: Adaptive Build

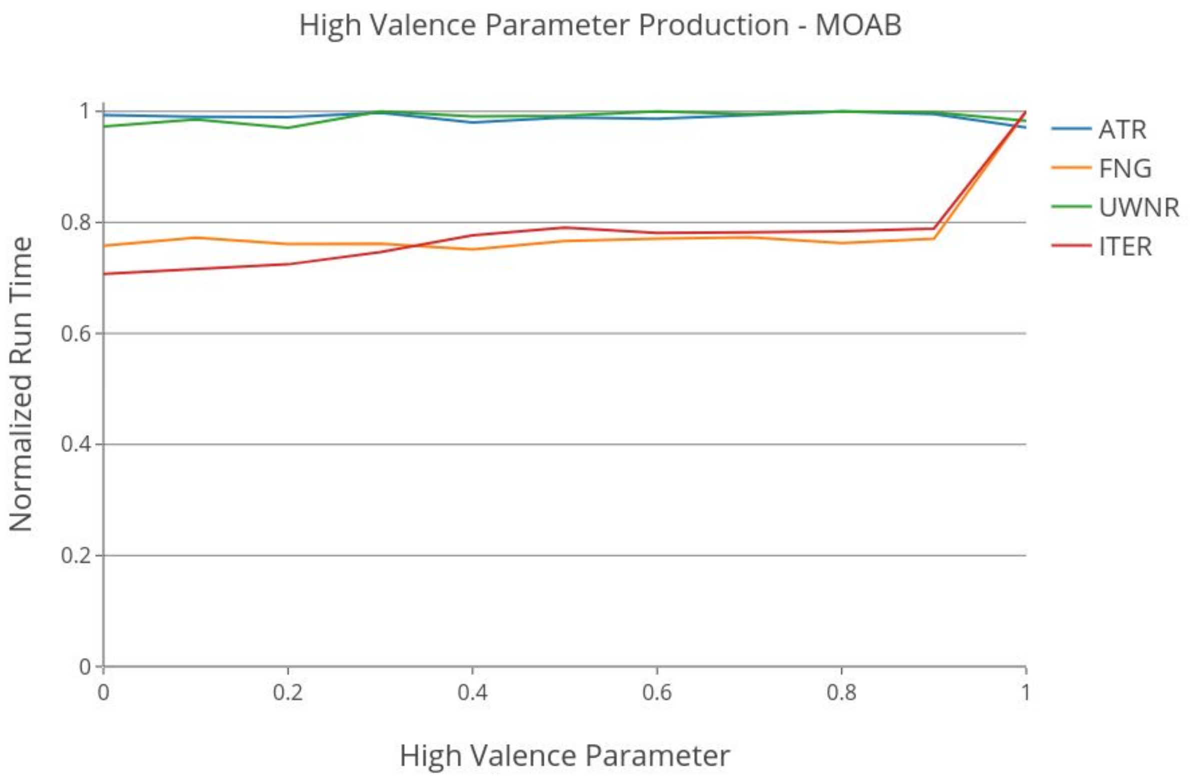


Adaptive Build Application: MOAB

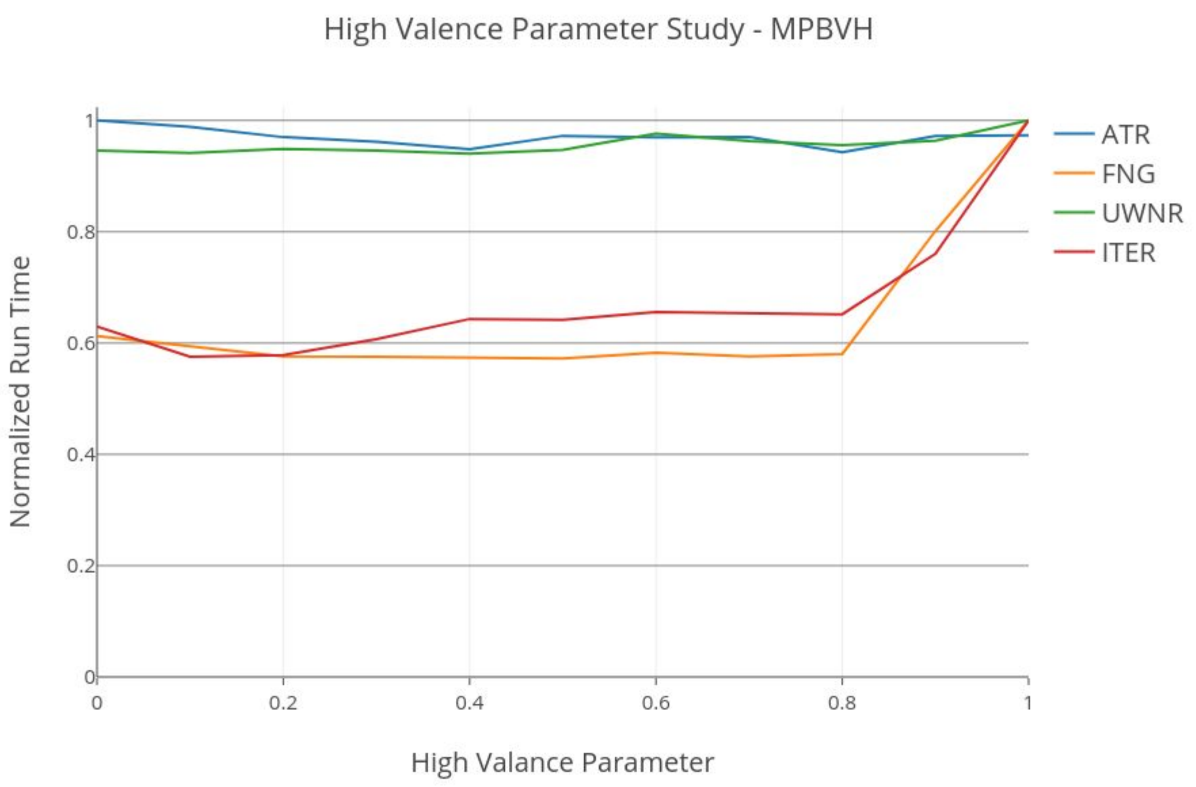
Model	HV Regions	HV Leaf Visit Rel. Freq. (%)	Run Time Relative to Unmodified BVH
FNG	514	3.83	0.76
ATR	836	0.07	0.97
UWNR	499	3.87	0.98
ITER	3522	7.34	0.72

HV Detection Parameter - 0.5

Adaptive Build Application: MOAB



Adaptive Build Application: MPBVH



Summary

- Adaptation of BVH construction for HV regions can have a significant impact on simulation run time.
  - The causes of ray tracing performance degradation may vary.
    - (large leaf nodes vs. overlapping nodes)
  - This may be specific to the tessellation algorithms in the DAGMC workflow.
  - HV regions are relatively straightforward to detect *provided a mesh database is available.*

Conclusion

Performance Impact Summary

Model	Method	Best improvement over unmodified DAG-MCNP
SNS	SDF	~1.00x
nTOF	SDF	1.41x
SHINE	SDF	1.56x
FNG	MPBVH	2.09x
UWNR	MPBVH	3.13x
ATR	MPBVH	4.56x
ITER	MPBVH	9.54x

Contributions

This work demonstrates improvements in performance for CAD-based MCRT ranging from factors of 2 to nearly 10, significantly reducing the computational cost of DAGMC simulations with no change in the numerical results.

The these methods and tools will significantly reduce time and cost associated with CAD-based MCRT analysis and are oriented toward future HPC architectures often used in this field.

Broader Impacts

- A predictive model for signed distance field utilization has been developed which in theory could also be used for other applications, such as domain decomposition for Woodcock delta tracking[12].
- The performance benefits demonstrated in DAGMC using the MPBVH can be leveraged in any application relying on MOAB.
- The use of OBBs to address high valence regions demonstrates their value for high valence mesh regions often found in engineering analysis tessellations.
- This work also provides evidence that use of a mesh framework to inform acceleration data structure construction can have a significant impact on simulation performance.

References

[1] Tautges, Timothy J., P. P. H. Wilson, et. al. 2009. Acceleration Techniques for Direct Use of CAD-Based Geometries in Monte Carlo Radiation Transport. In International Conference on Mathematics, Computational Methods & Reactor Physics (M&C 2009). Saratoga Springs, NY: American Nuclear Society.

[2] Trelis user's guide. https://www.csimsoft.com/help/trelishelp.html. Accessed: 2018-05-09.

[3] Tautges, T. J., R. Meyers, K. Merkley, C. Stimpson, and C. Ernst. 2004. MOAB: A Mesh-Oriented Database. SAND2004-1592, Sandia National Laboratories. Report.

[4] Smith, Brandon M. 2011. Robust Tracking and Advanced Geometry for Monte Carlo Radiation Transport. PhD Nuclear Engineering and Engineering Physics, University of Wisconsin-Madison, Madison, WI,United States.

[5] TEAM, X-5 MONTE CARLO. 1987. MCNP – a general monte carlo n-particle transport code, version 5. Tech. Rep. LA-UR-03-1987, Los Alamos National Laboratory.

[6] V-ray Theory Guide. https://docs.chaosgroup.com/display/VR2R/Basic+Ray+Tracing. Accessed: 2018-06-03.

[7] Wald, Ingo, Sven Woop, Carsten Benthin, Gregory S. Johnson, and Manfred Ernst. 2014. Embree: A kernel framework for efficient CPU ray tracing.ACM Trans. Graph.33(4):143:1–143:8.

References

[8] Vaidyanathan, K., T. Akenine-Möller, and M. Salvi. 2016. Watertight ray traversal with reduced precision. In Proceedings of high performance graphics, 33–40. HPG ’16, Aire-la-Ville, Switzerland, Switzerland: Eurographics Association.

[9] Gottschalk, Stefan, Ming C Lin, and Dinesh Manocha. 1996. Obbtree: A hierarchical structure for rapid interference detection. In Proceedings of the 23rd annual conference on computer graphics and interactive techniques, 171–180. ACM.

[10] 2008. IEEE standard for floating-point arithmetic. IEEE Std. 754-20081–70.

[11] Platis, Nikos, and Theoharis Theoharis. 2003. Fast Ray-Tetrahedron Intersection Using Plucker Coordinates. Journal of Graphics Tools 8(4):37–48.

[12] Leppänen, Jaakko. 2010. Performance of woodcock delta-tracking in lattice physics applications using the serpent monte carlo reactor physics burnup calculation code. Annals of Nuclear Energy 37(5):715 –722.

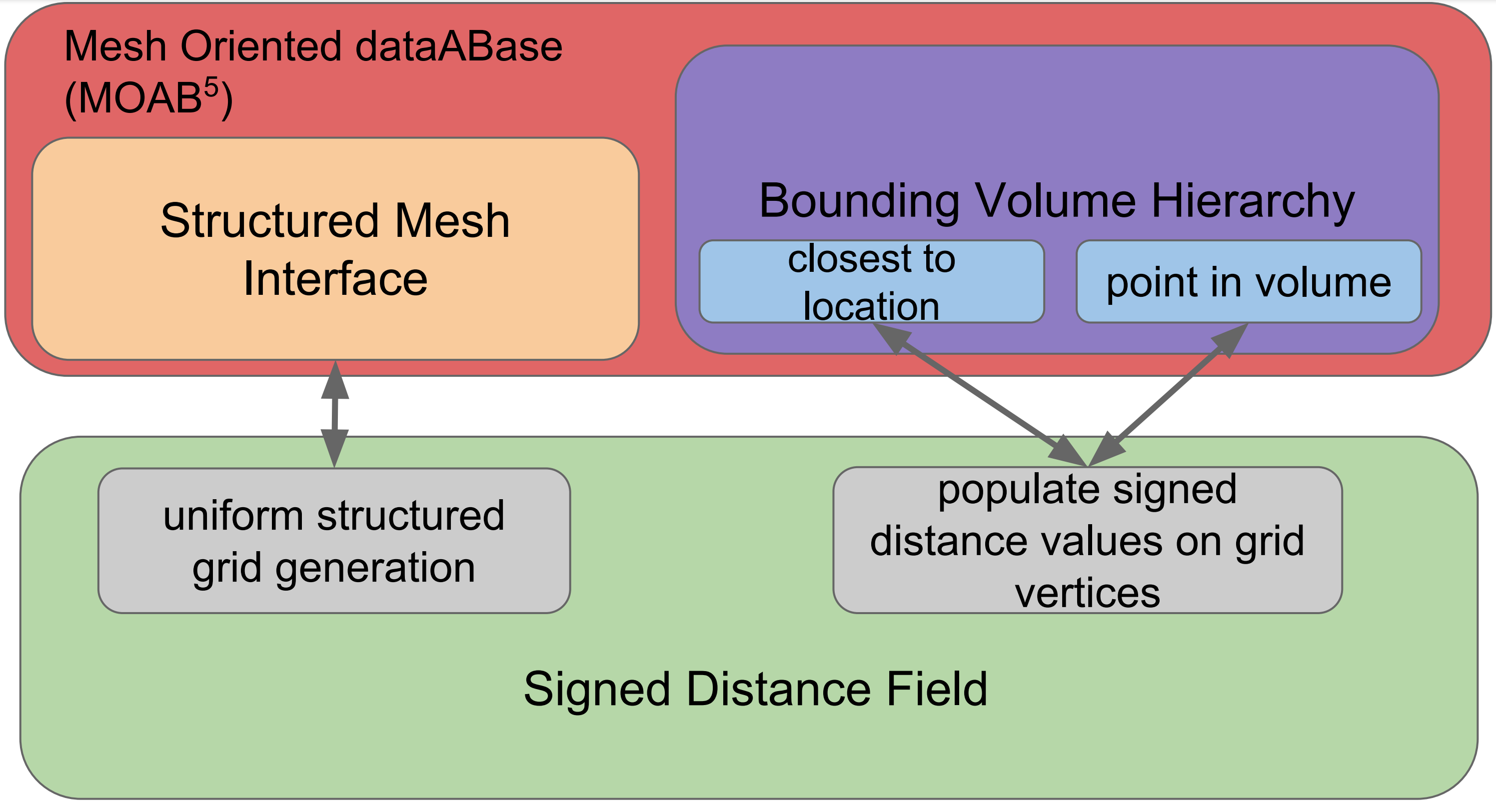
Questions?

CNERG

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

U.S.NRC

SDF Implementation



Modeling Parameters

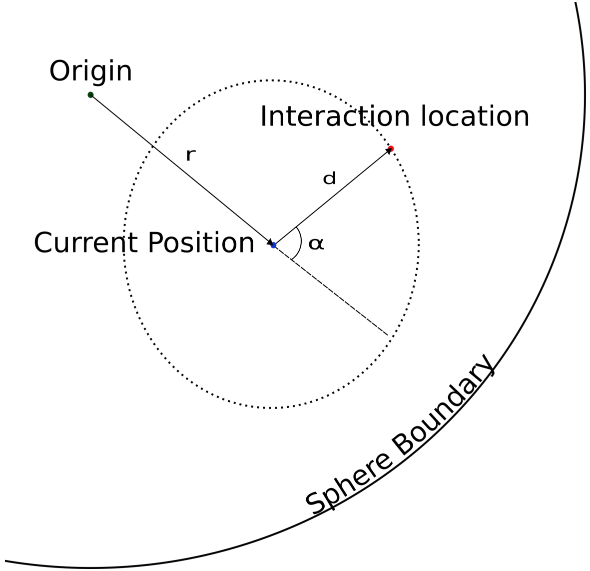
Preconditioner Utilization Condition:

FIX

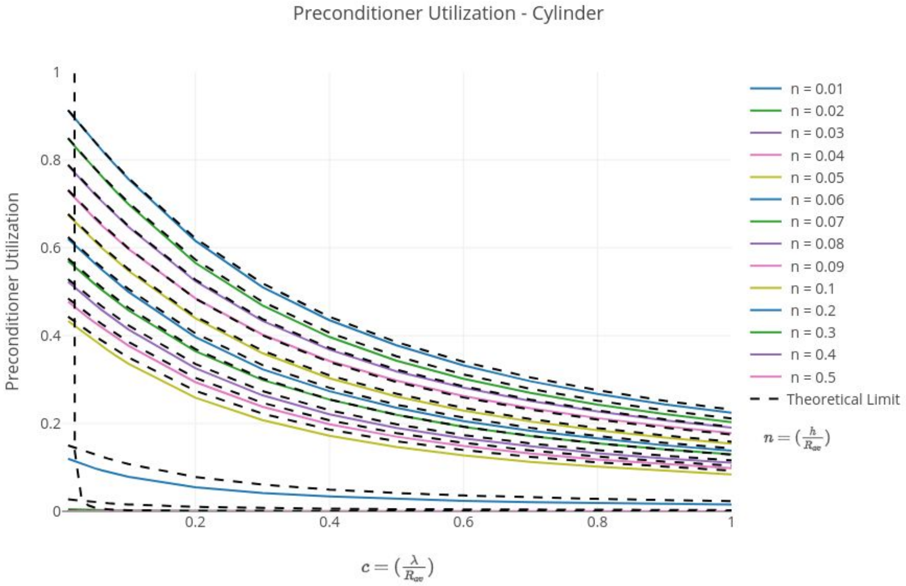
$$R - r + R - |\vec{n}(d, \alpha)| > d$$

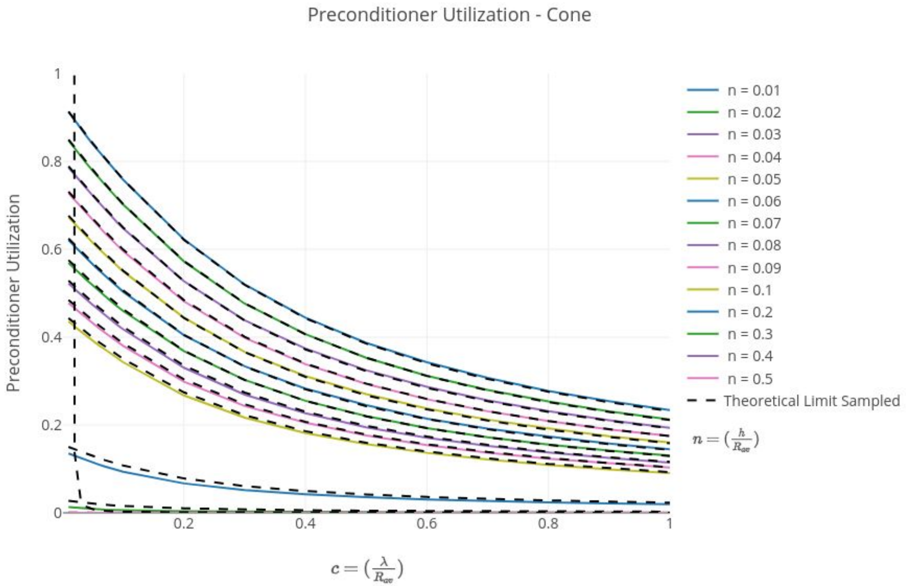
Limit on α for particle utilization

$$\alpha_{min} > \arccos\left(\frac{(2R - r - d)^2 - d^2 - r^2}{2dr}\right)$$



Utilization Modeling - Other Geometries

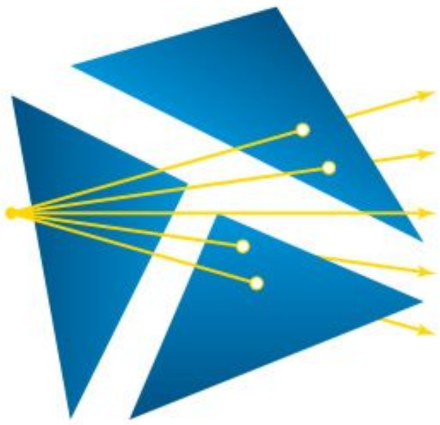




SIMD Ray Tracing and Monte Carlo

Two approaches to using SIMD in ray tracing:

- N:1 - test many rays against a single box at once
- 1:N - test one ray against many boxes at once

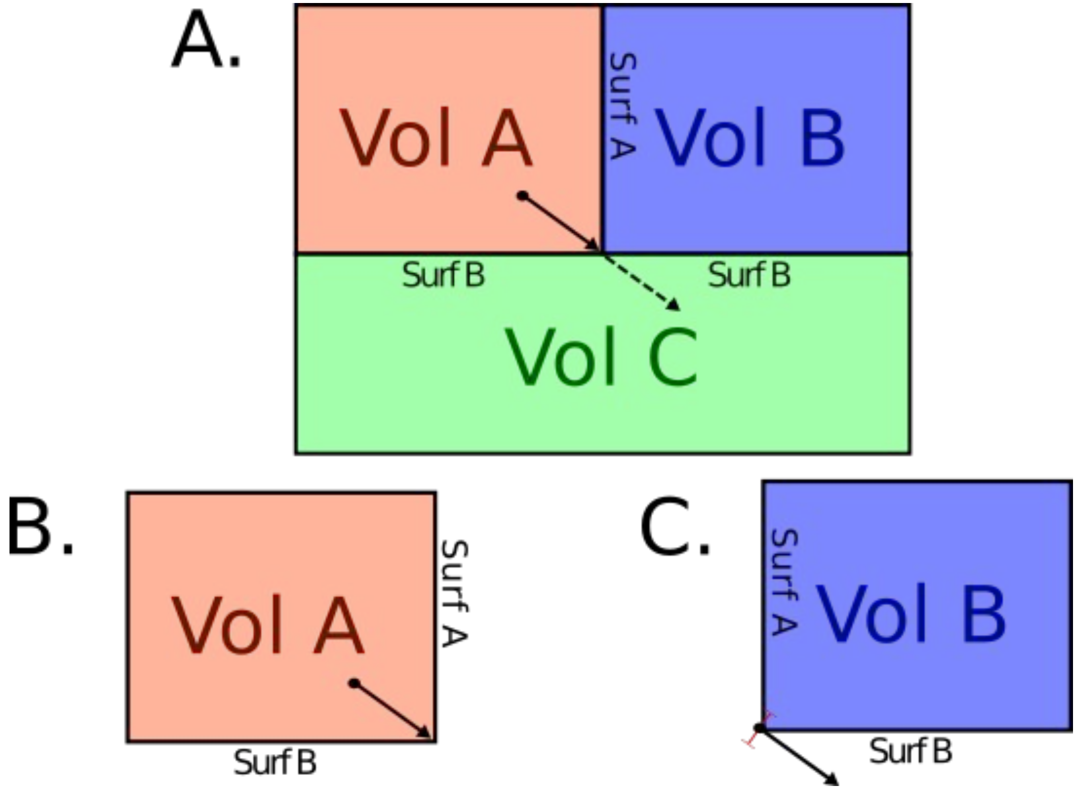


Coherent

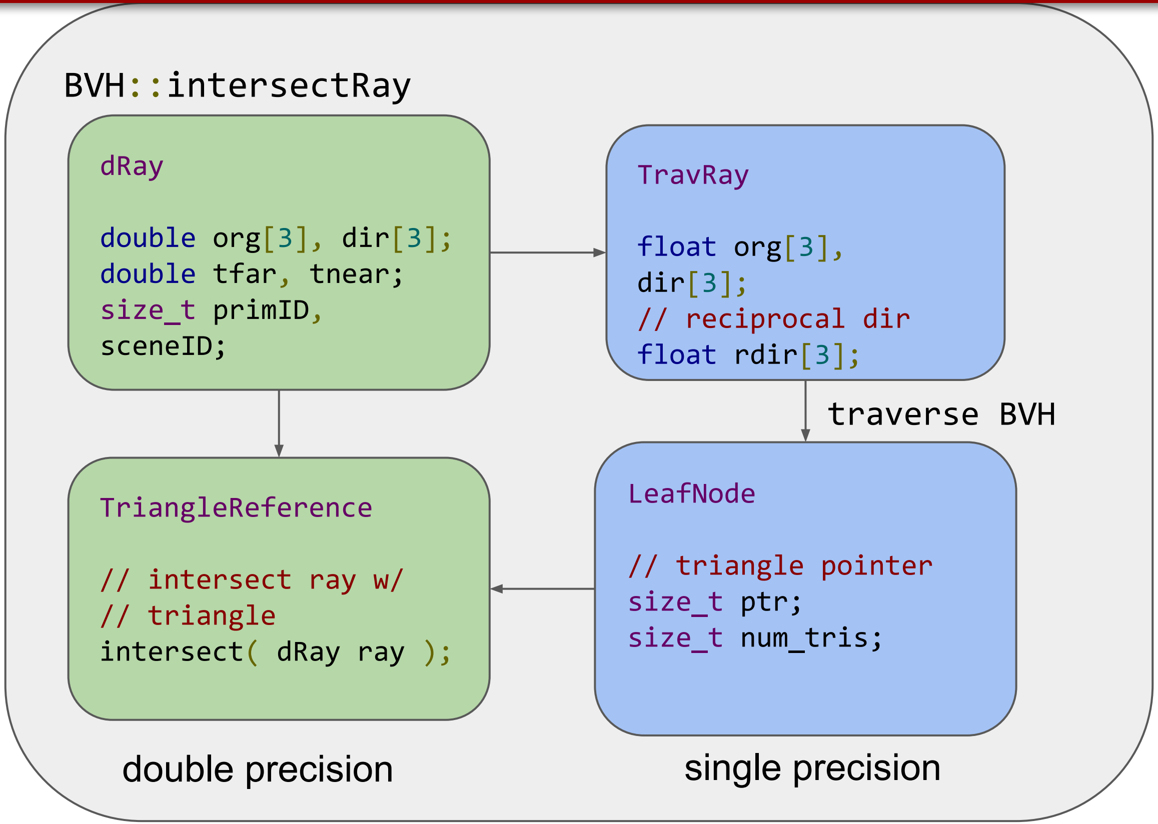
Incoherent

* image courtesy of Intel

EmDAG Lost Particles



Implementation



Surface Reflection