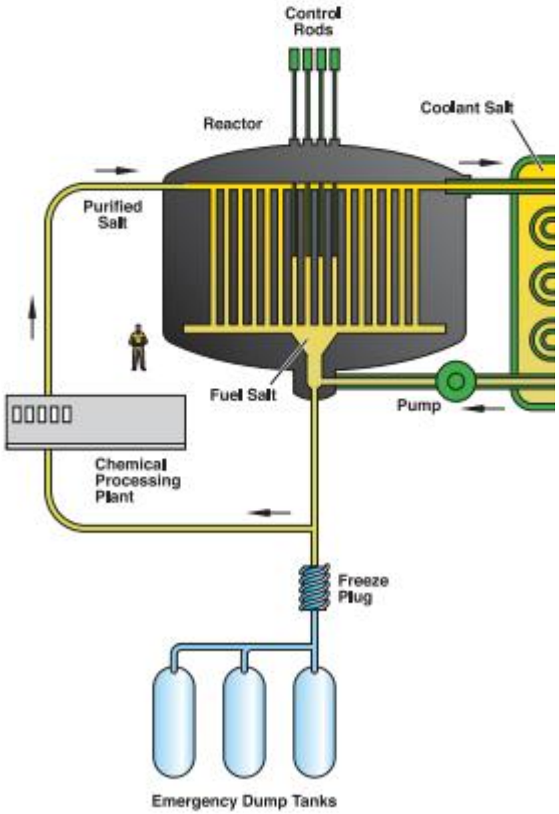
# Exascaling Nuclear Innovation

Prof. Rachel Slaybaugh
4 March, 2016
Conference on Data Analysis
Santa Fe, NM

Berkeley
UNIVERSITY OF CALIFORNIA

# Nuclear Innovation is Needed

- Global climate and environment goals
- Public health and economic prosperity
- Global and domestic security goals



Enable New Reactors

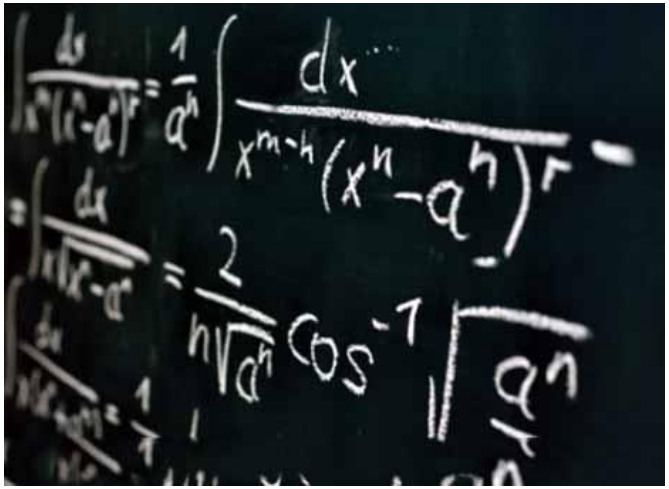
Enhance Nuclear Security

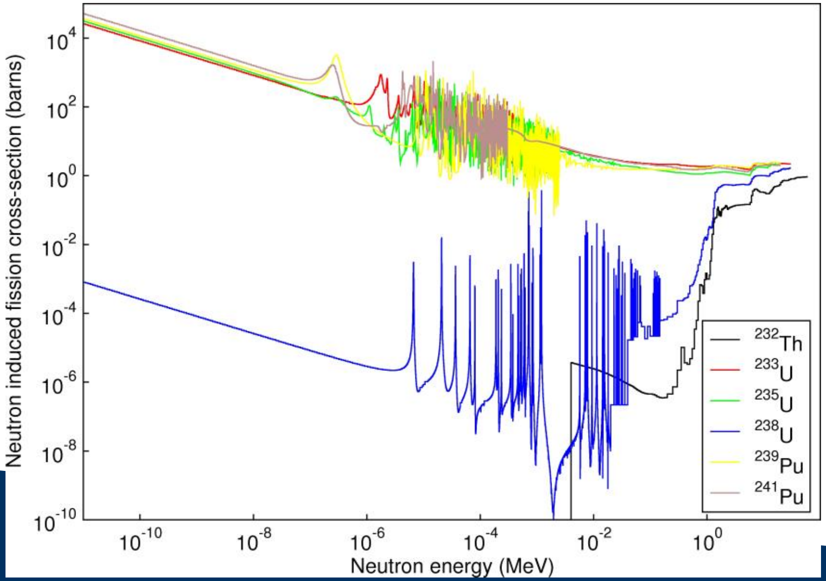
Develop Medical Devices

# my contribution: numerical methods

To facilitate innovation,
we need *predictive simulation*

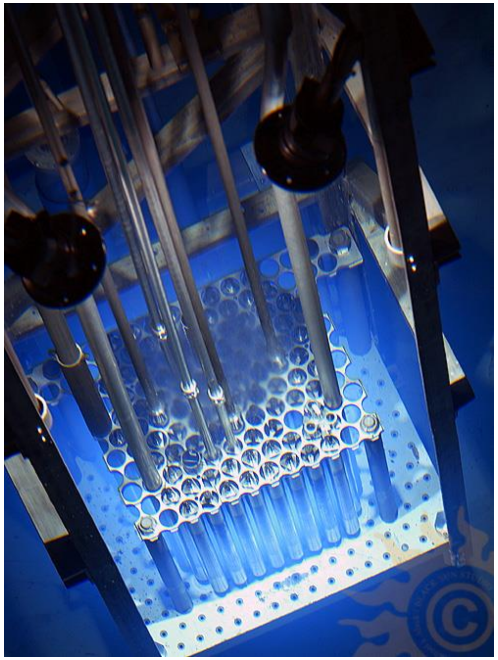
- I build tools (translate applied math into code) used to design and analyze these systems
- I focus on high performance computing
- and inform algorithm development with physics of problems of interest





# outline

- Background
  - What exactly are you solving? Why is it hard? What is the current state?
- Algorithms for advanced architectures
  - Big and heterogeneous machines
  - + physics provide opportunities
- Nuclear data
  - Yes, there are problems
  - But we can make them better
- Processing data while computing?



Berkeley
UNIVERSITY OF CALIFORNIA

# finding all of the neutrons

$$[\hat{\Omega}\cdot\nabla+\Sigma(\vec{r},E)]\psi(\vec{r},\hat{\Omega},E)=$$
$$\int dE'\int d\hat{\Omega}'\,\Sigma_s(\vec{r},E'\rightarrow E,\hat{\Omega}'\cdot\hat{\Omega})\psi(\vec{r},\hat{\Omega}',E')$$
$$+\frac{\chi(E)}{k}\int dE'\,\nu\Sigma_f(\vec{r},E')\int d\hat{\Omega}'\,\psi(\vec{r},\hat{\Omega}',E')$$

- $\psi(\vec{r},\widehat{\Omega},E)$: eigenvector; angular neutron flux (n / cm^2-steradian)

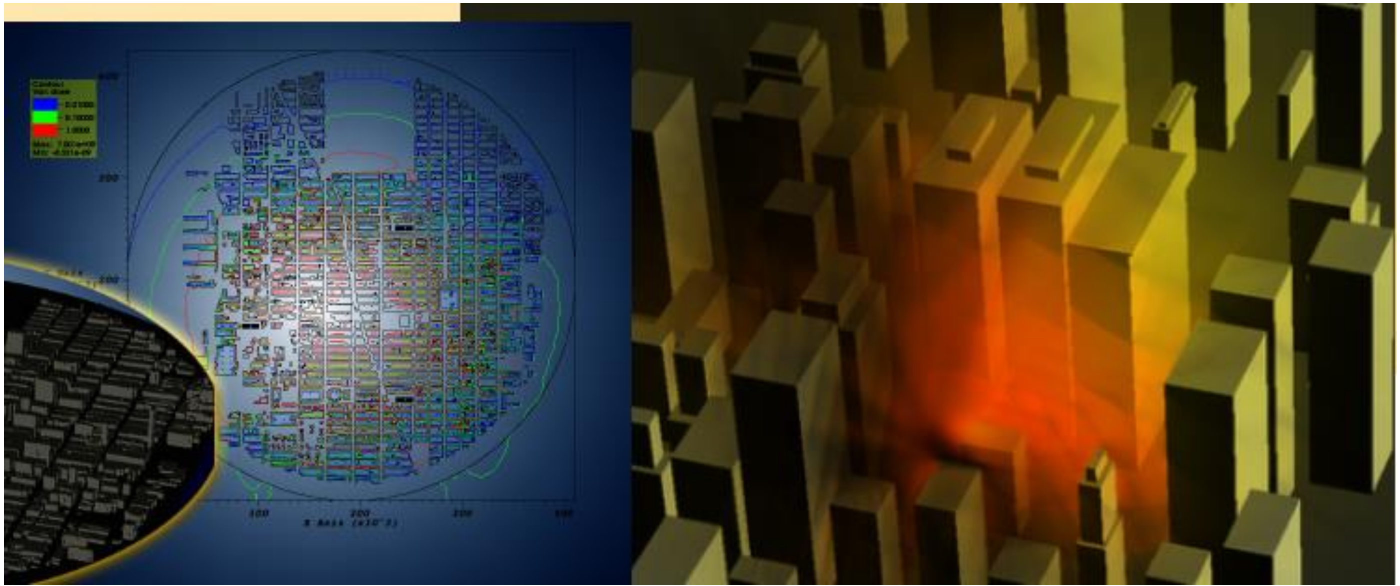
- *k:* dominant eigenvalue; governs steady-state system behavior

- $\Sigma(\vec{r},E)$: probability of neutrons interacting with a material (cm^-1)

Berkeley
UNIVERSITY OF CALIFORNIA

# accurately is hard

- 6-D phase space: location (3), direction (2), energy (1)
- Can be geometrically large and/or physically complex and/or coupled to other physics
- The physics data is *Complicated*



- Strategies:

Deterministic

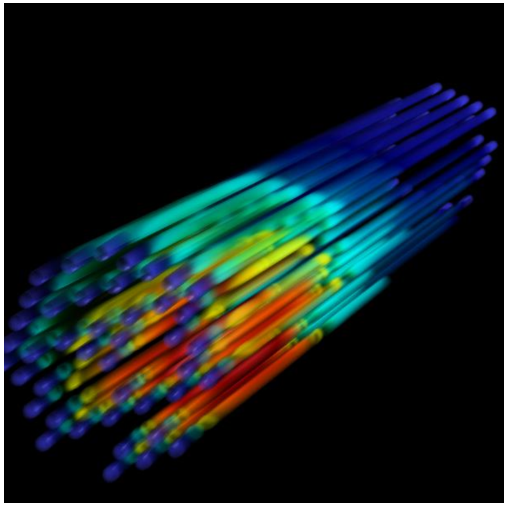
Monte Carlo

# what is the current state?

- Existing tools can predict what we already know, not new things
  - We need increased accuracy, resolution, turnaround time, and confidence in our solutions
- We need better nuclear data
  - Increasing resolution is creating a new garbage in problem…
- We need faster analysis strategies
  - We can generate lots of answers, but it is taking increasing amounts of time and storage to process them

Exascale computing to the rescue?

# algorithms = physics + architecture

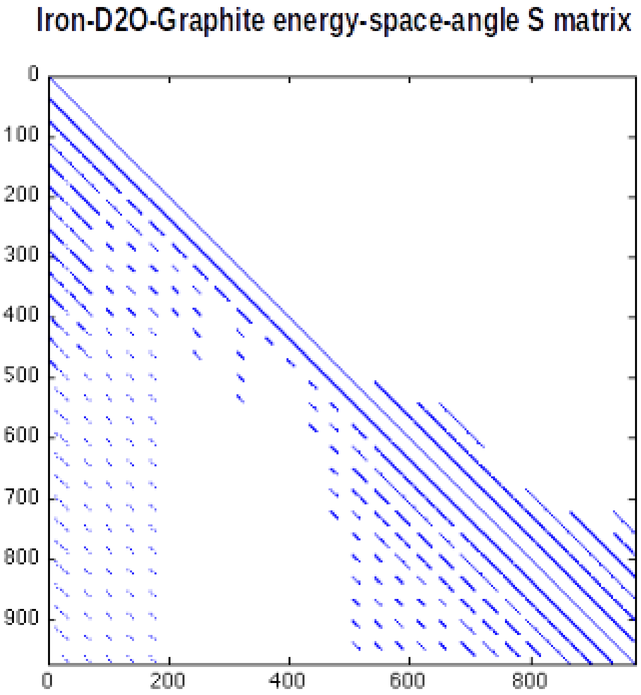


# quick brief: deterministic methods

- Discretize phase space; represent as a matrix; apply linear algebra-type solvers
  - Space: mesh (think finite difference/volume)
  - Energy: break into multiple groups
  - Angle: capture with a quadrature set
- Quality of solution is tied to
  - Quality of discretization
  - Accuracy of solution methods
- Can be memory and FLOP intensive
- Strategically-designed parallelization techniques

# ex. 1: block Krylov + RQI + MGE

- With deterministic methods, our matrices look like

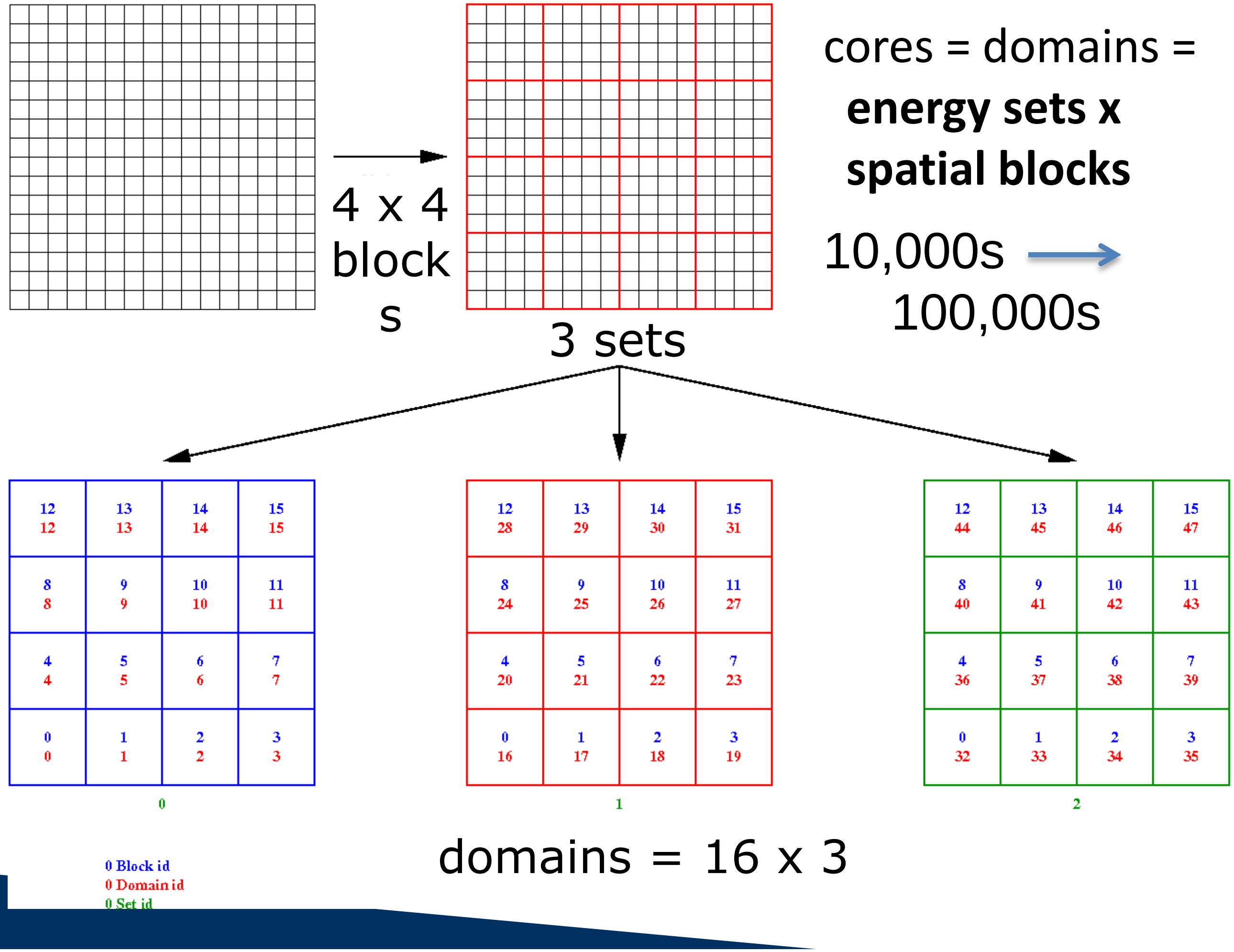


- Historical solution:
  - Inner iterations solve space and angle for each energy group (Krylov over one group)
  - Outer iterations over energy (Gauss Seidel)

- Issues:
  - GS is serial in energy; slow with upscattering

# larger memory: block Krylov

- Machines can store multi-group-sized iteration vectors
- Krylov over all groups: *space, angle, and energy* ***together***
- Better convergence
- Energy parallelization



cores = domains = **energy sets x spatial blocks**

10,000s → 100,000s

| | | | |
|---|---|---|---|
| 12<br>12 | 13<br>13 | 14<br>14 | 15<br>15 |
| 8<br>8 | 9<br>9 | 10<br>10 | 11<br>11 |
| 4<br>4 | 5<br>5 | 6<br>6 | 7<br>7 |
| 0<br>0 | 1<br>1 | 2<br>2 | 3<br>3 |

0

| | | | |
|---|---|---|---|
| 12<br>28 | 13<br>29 | 14<br>30 | 15<br>31 |
| 8<br>24 | 9<br>25 | 10<br>26 | 11<br>27 |
| 4<br>20 | 5<br>21 | 6<br>22 | 7<br>23 |
| 0<br>16 | 1<br>17 | 2<br>18 | 3<br>19 |

1

| | | | |
|---|---|---|---|
| 12<br>44 | 13<br>45 | 14<br>46 | 15<br>47 |
| 8<br>40 | 9<br>41 | 10<br>42 | 11<br>43 |
| 4<br>36 | 5<br>37 | 6<br>38 | 7<br>39 |
| 0<br>32 | 1<br>33 | 2<br>34 | 3<br>35 |

2

0 Block id
0 Domain id
0 Set id

domains = 16 x 3

G.G. Davidson et al., “Massively Parallel, Three-Dimensional Transport Solutions for the k-Eigenvalue Problem,” *Nuc. Sci. and Eng*. **177** (2014) 111-125.

# less time, fewer iterations

- Block Krylov enabled two other new methods as well (RQI, MGE)
- Applied to problem with 1.73 trillion unknown on ~140,000 cores

| Solver | Precond | Krylov | Eigen | Time (m) |
|---|---|---|---|---|
| Original | none | 5,602 | 149 | 612.2 |
| New | w1r2v2 | 70 | 5 | 54.8 |

- 10x improvement
- Also strong scale very well

R.N. Slaybaugh, T.M. Evans, G.G. Davidson, and P.P.H. Wilson, "Rayleigh Quotient Iteration with a Multigrid in Energy Preconditioner for Massively Parallel Neutron Transport," Proceedings of M&C-SNA in Nashville, TN, April 2015.

# quick brief: Monte Carlo

- Physics expressed continuously; sampled with random numbers

- Quality of solution is tied to
  - Number of samples in phase space
  - Adequacy of sampling phase space
- *Variance Reduction* methods can reduce variance faster while maintaining a fair game

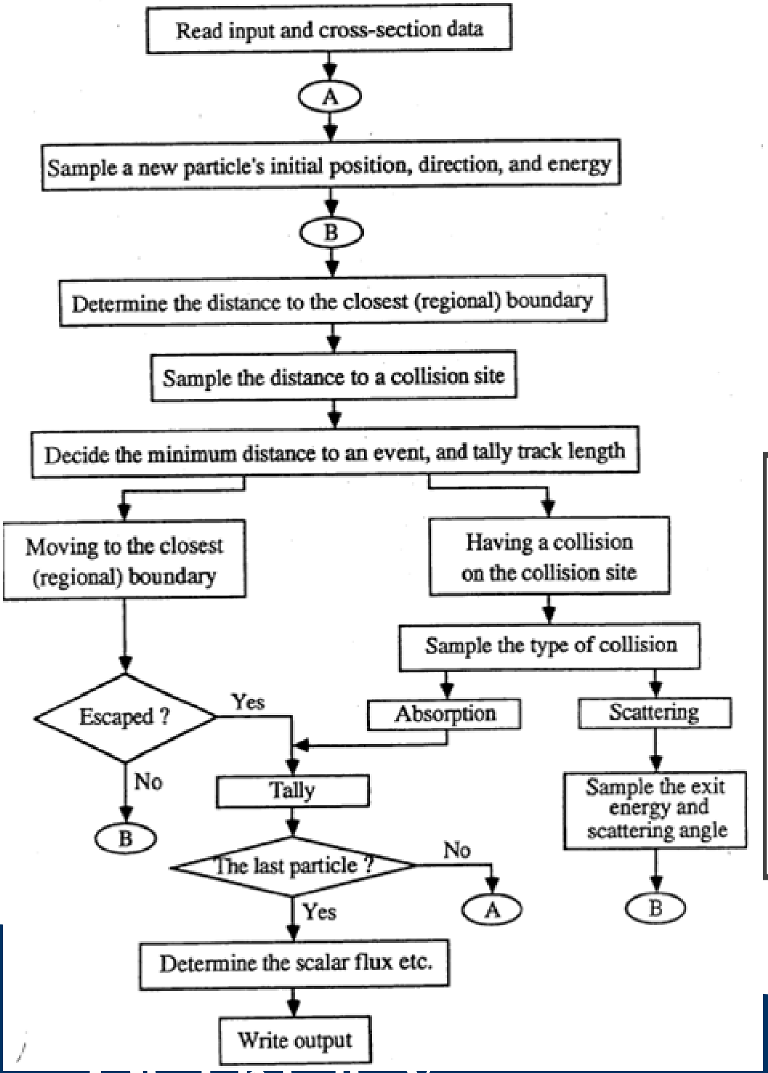
- Can be memory and FLOP intensive
- Historically straightforward to parallelize

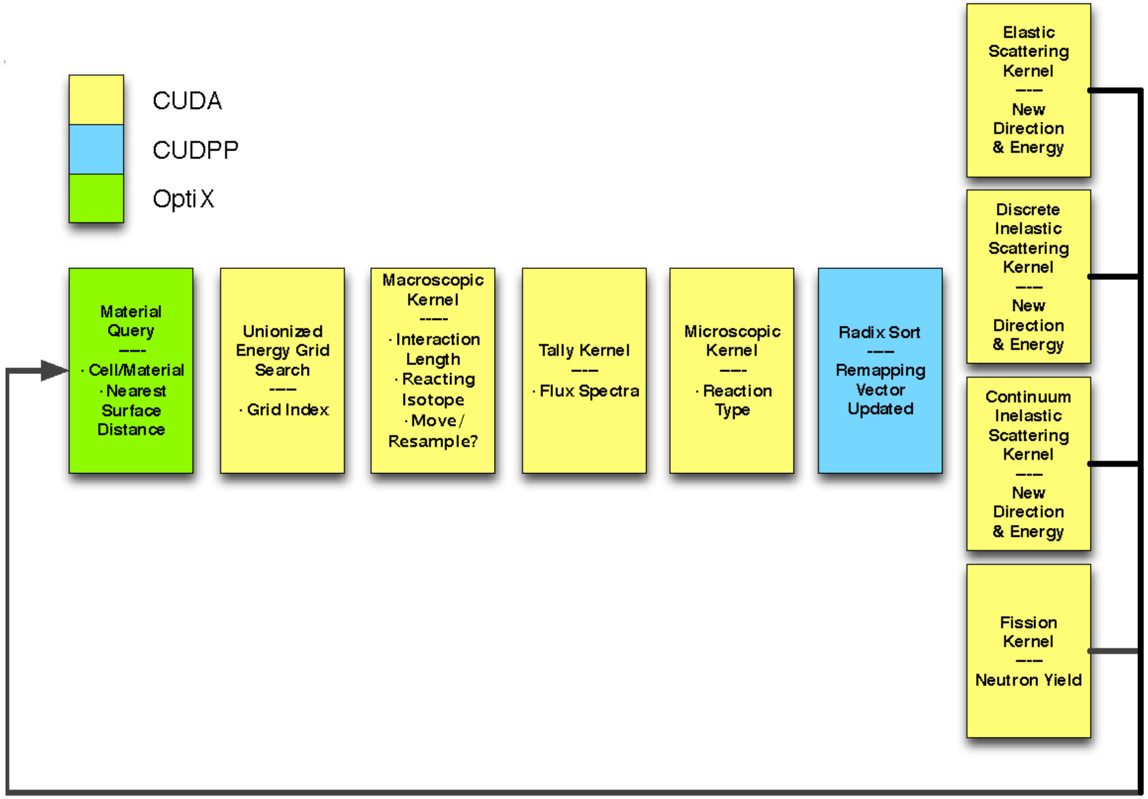
# ex. 2: Monte Carlo on GPUs

- WARP: Weaving All the Random Particles
- 3D continuous-energy Monte Carlo neutron transport code developed for efficient implementation of the algorithm on GPUs
- Relative to CPUs, GPUs have higher aggregate memory bandwidth, much higher floating-point operations per second (FLOPS), lower energy consumption per FLOP
- CPU-optimized parallel algorithms not directly portable to GPUs
- Particle transport codes need to be rewritten to execute efficiently on GPUs

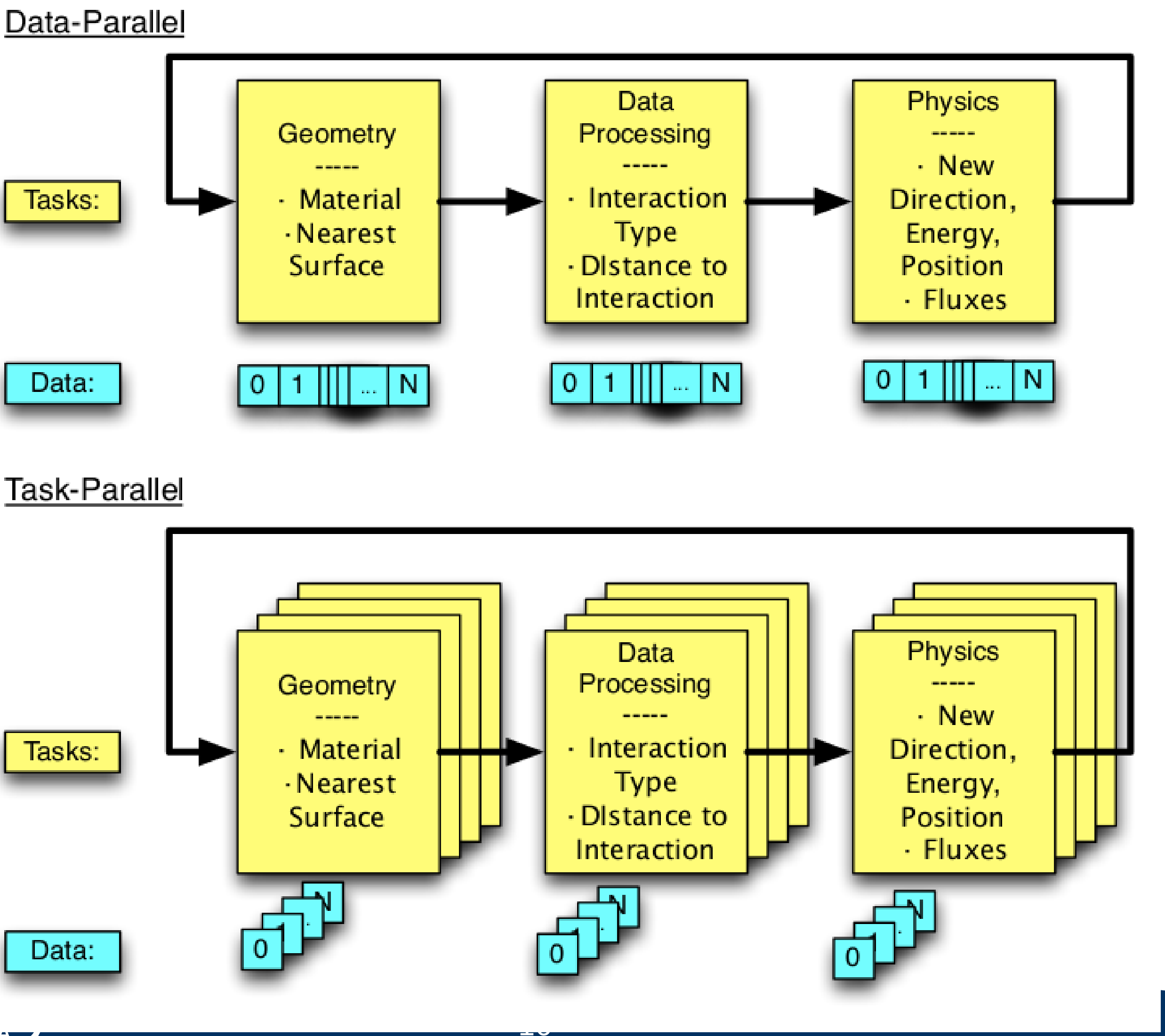
Ryan  M. Bergmann, Jasima L. Vujic. "WARP – A framework for continuous energy Monte Carlo neutron transport in general 3D geometries on GPUs," *Annals of Nuclear Energy* **77** (2015) 176-193.

# WARP uses an event-based algorithm


Read input and cross-section data
A
Sample a new particle's initial position, direction, and energy
B
Determine the distance to the closest (regional) boundary
Sample the distance to a collision site
Decide the minimum distance to an event, and tally track length
Moving to the closest (regional) boundary
Having a collision on the collision site
Sample the type of collision
Escaped ?
Yes
No
B
Absorption
Scattering
Tally
Sample the exit energy and scattering angle
The last particle ?
No
A
Yes
B
Determine the scalar flux etc.
Write output



# and is data parallel



Data-Parallel

Tasks:

Geometry
-----
· Material
· Nearest Surface

Data Processing
-----
· Interaction Type
· DIstance to Interaction

Physics
-----
· New Direction, Energy, Position
· Fluxes

Data:

0 1 ... N

0 1 ... N

0 1 ... N

Task-Parallel

Tasks:

Geometry
-----
· Material
· Nearest Surface

Data Processing
-----
· Interaction Type
· DIstance to Interaction

Physics
-----
· New Direction, Energy, Position
· Fluxes

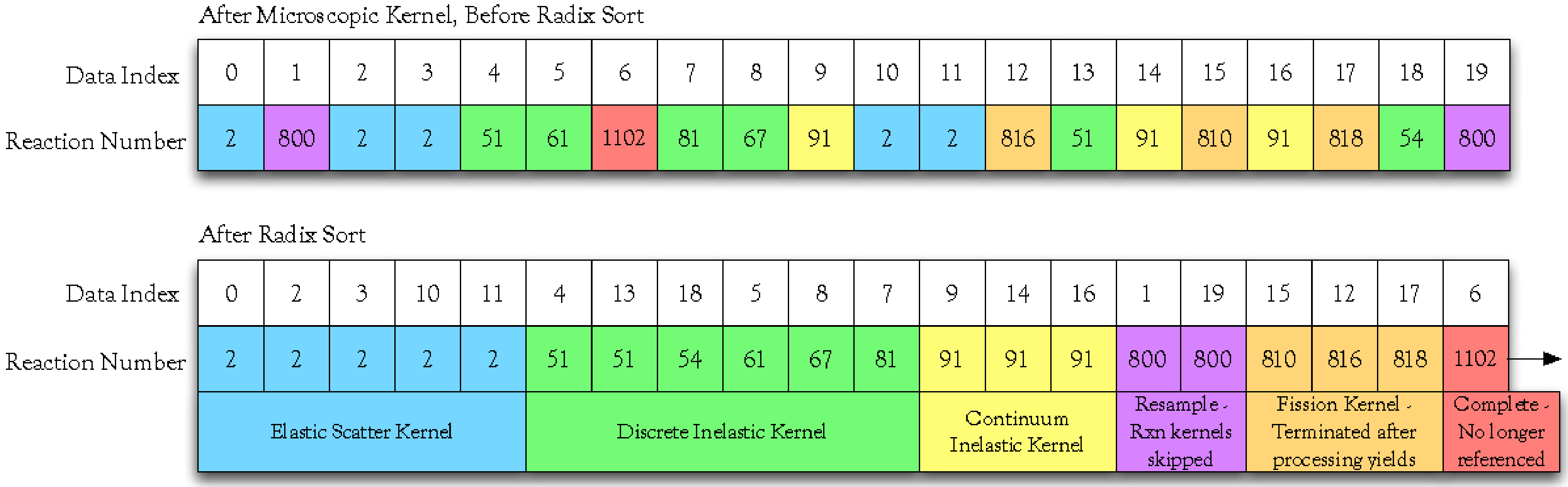
Data:

0 N

0 N

0 N

# WARP remaps data



After Microscopic Kernel, Before Radix Sort

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data Index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| Reaction Number | 2 | 800 | 2 | 2 | 51 | 61 | 1102 | 81 | 67 | 91 | 2 | 2 | 816 | 51 | 91 | 810 | 91 | 818 | 54 | 800 |

After Radix Sort

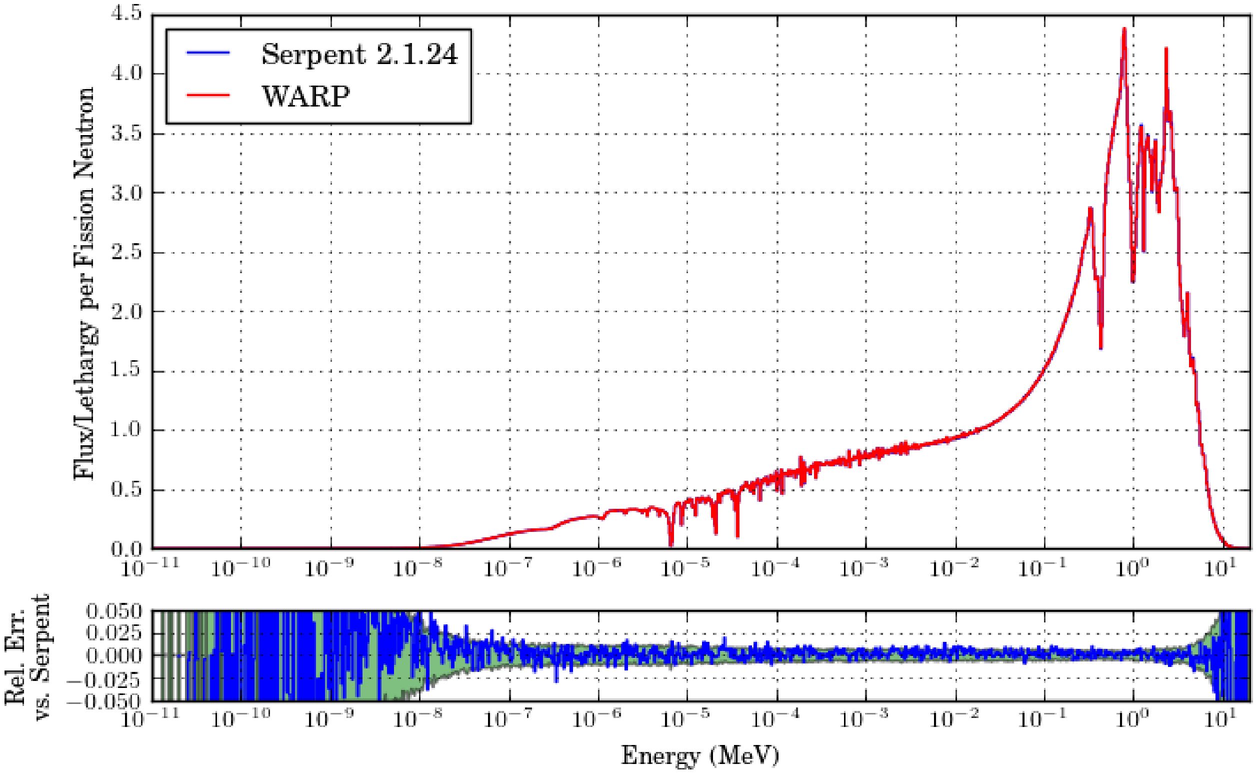
| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data Index | 0 | 2 | 3 | 10 | 11 | 4 | 13 | 18 | 5 | 8 | 7 | 9 | 14 | 16 | 1 | 19 | 15 | 12 | 17 | 6 |
| Reaction Number | 2 | 2 | 2 | 2 | 2 | 51 | 51 | 54 | 61 | 67 | 81 | 91 | 91 | 91 | 800 | 800 | 810 | 816 | 818 | 1102 |
| | Elastic Scatter Kernel | | | | | Discrete Inelastic Kernel | | | | | | Continuum Inelastic Kernel | | | Resample - Rxn kernels skipped | | Fission Kernel - Terminated after processing yields | | | Complete - No longer referenced |

A radix key-value sort is used to create a remapping vector, grouping neutrons by reaction in order to minimize thread divergence

# results match standard solvers

WARP took 2.0615 s
(k20; 32 threads)

Serpent took 137 s
(AMD Opteron 6172; 12 cores)

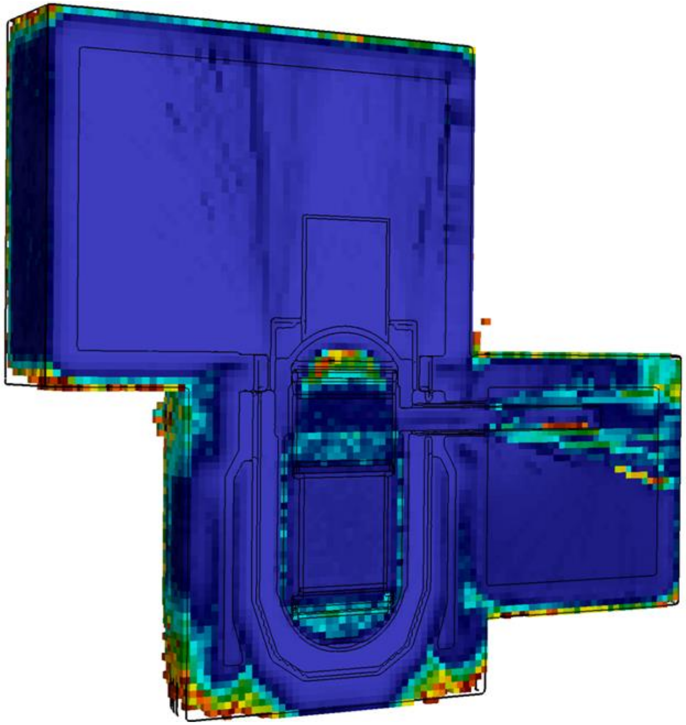


# quick brief: hybrid methods

- Monte Carlo highly accurate; can be slow
- Deterministic methods usually fast; can have accuracy issues
- Hybrid methods use deterministic solutions to create variance reduction parameters for MC
  - Particles are assigned weights that map to impact
  - Set how to update weights
  - Set how to bias the source

# ex. 3: angle-informed methods

- Radiation shielding is a tough problem, especially when there are strong angular anisotropies

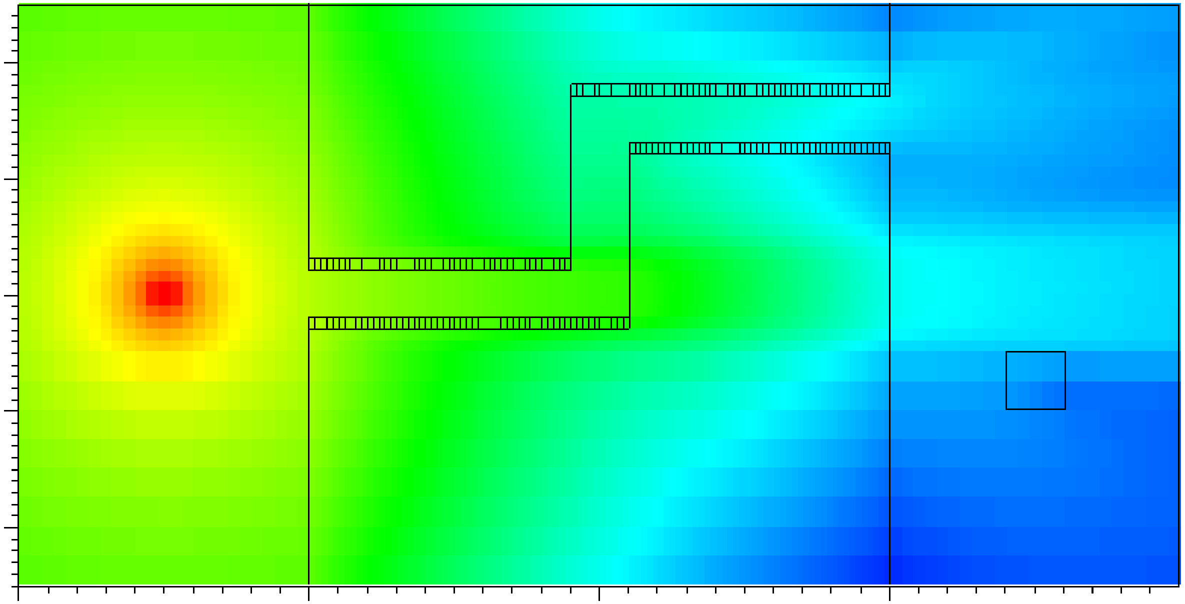


- Hybrid methods currently only include space and energy information
- Including angle explicitly is too costly
- Other attempts haven't worked well
- But we are having trouble…

Berkeley
UNIVERSITY OF CALIFORNIA

# start with importance

- Many hybrid methods use adjoint transport information, which we think of as the importance of a source particle to the solution



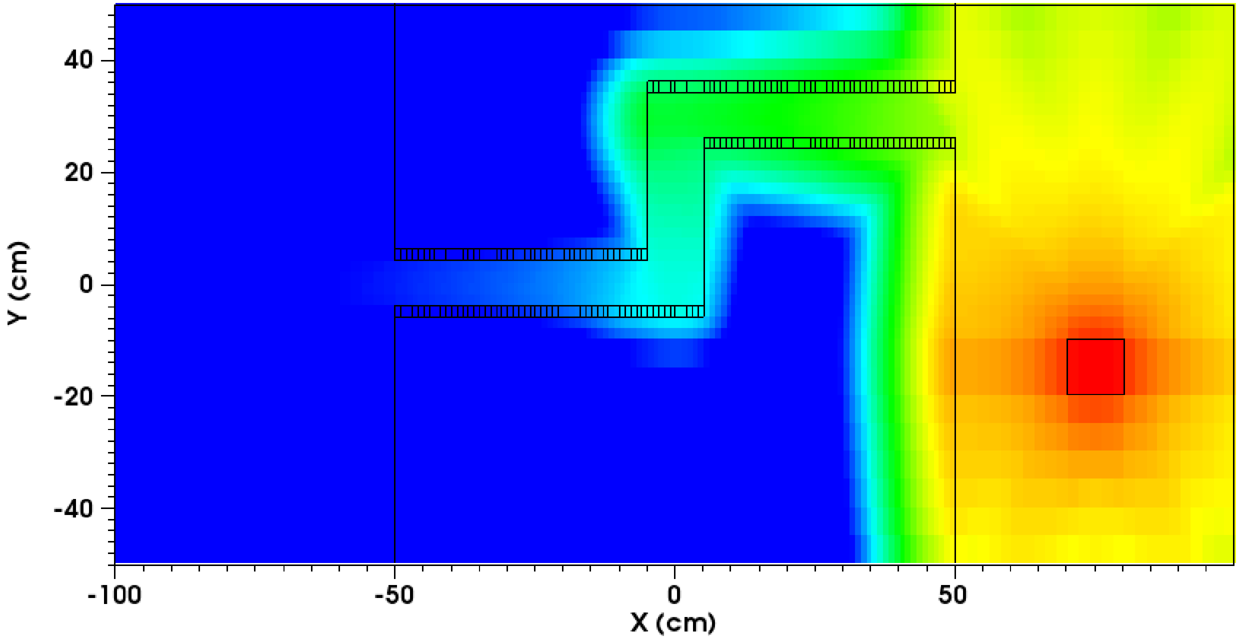
Neutrons in the forward problem will flow from the source to the detector..

# start with importance

- Many hybrid methods use adjoint transport information, which we think of as the importance of a source particle to the solution

Adjoint particles are how each part of phase space contributes to the solution:

importance map



# and a good method

- The state of the art in current space-energy methods is called FW-CADIS

- It uses importance information that is only a function of space and energy

$$\phi^{+}(\vec{r},E)=\int\psi^{+}(\vec{r},E,\hat{\Omega})d\hat{\Omega}$$

to set VR parameters

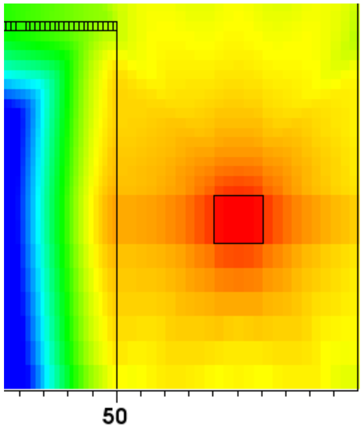
- It has been successful in many problems

# and modify it to include angle

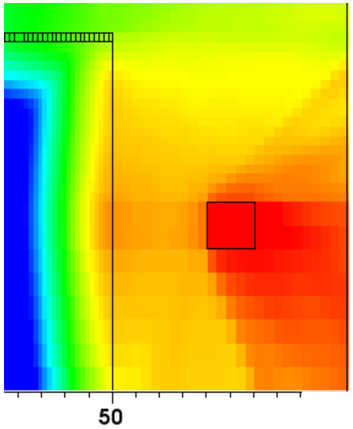
- We've modified FW/CADIS to include angular information without changing the method's end use

$$\phi^+(\vec{r},E) = \frac{\int \psi^+(\vec{r},E,\hat{\Omega})\psi(\vec{r},E,\hat{\Omega})d\hat{\Omega}}{\int \psi(\vec{r},E,\hat{\Omega})d\hat{\Omega}}$$

- Enabled by ability to write and store $\psi^+(\vec{r},E,\hat{\Omega})$

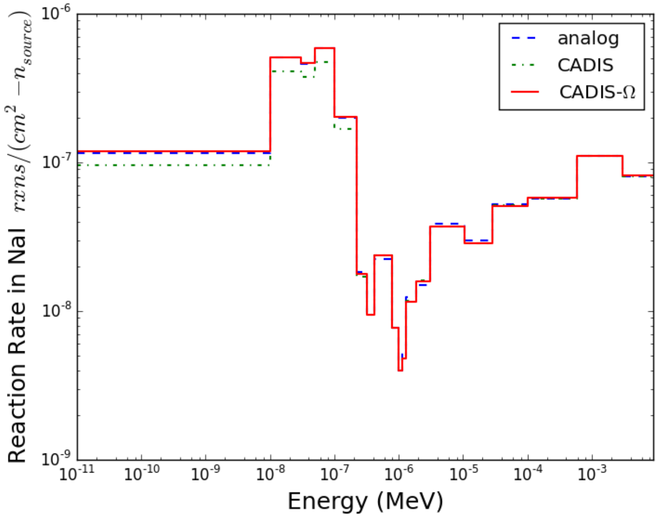


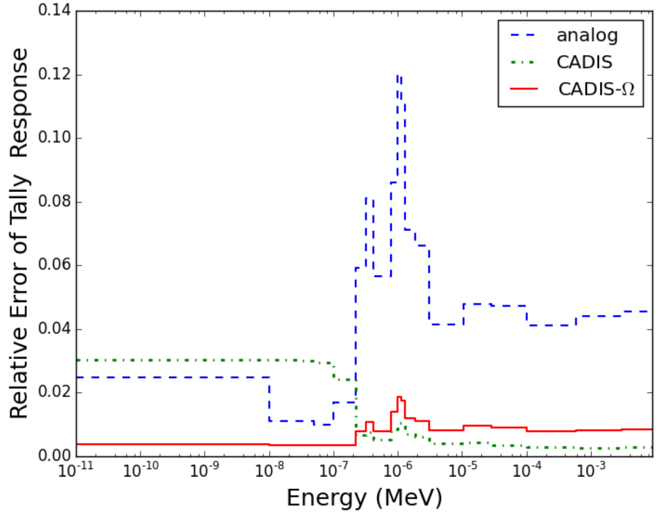
Original Method



Our Method

# we have good initial results





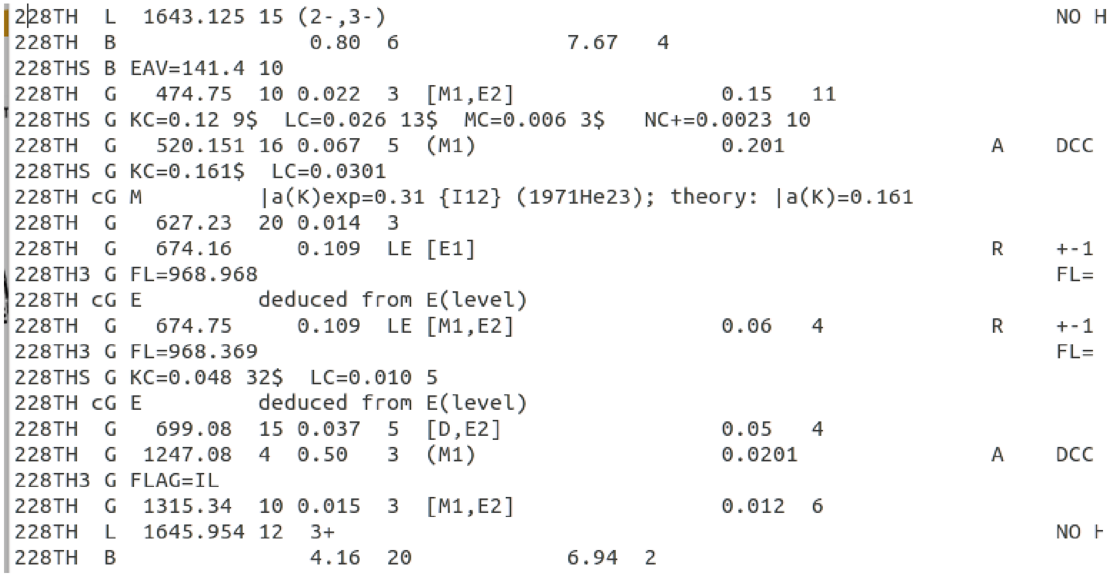
| | $FOM_{MC}$ | $FOM_{Adjusted}$ |
|---|---|---|
| CADIS | 8 | 7 |
| CADIS-Ω | 466 | 354 |

Adjusted CADIS has:

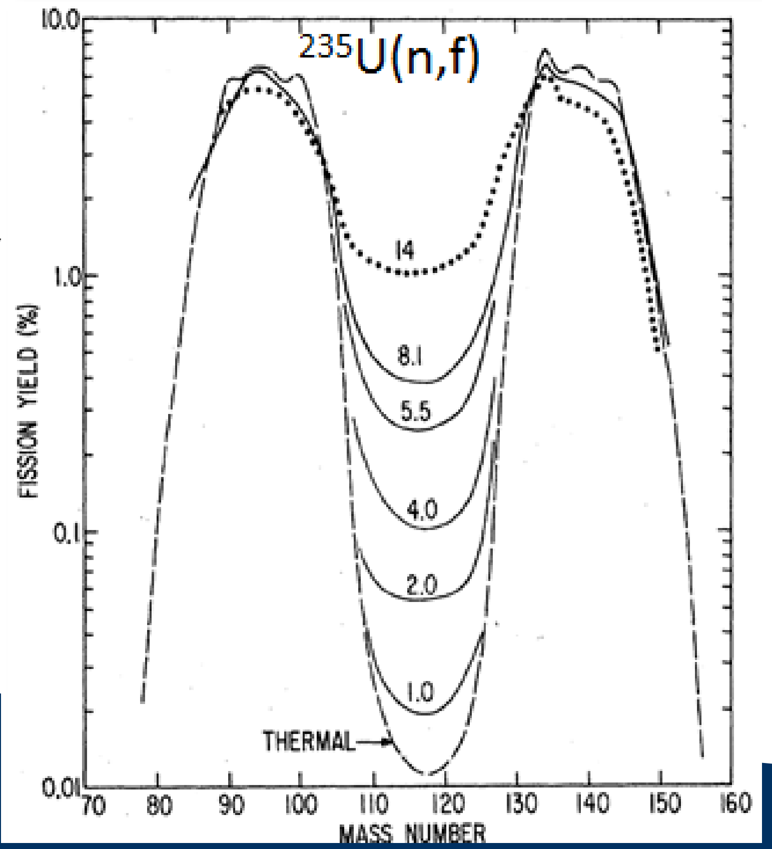
- A relatively uniform uncertainty distribution
- much faster runtimes than CADIS
- better overall results (comparing all energies)

M. Munk, et al., “An Angle-Informed Hybrid Method for CADIS and FW-CADIS.” Proceedings of the PHYSOR 2016 Meeting in Sun Valley, ID, May 2016. (accepted).

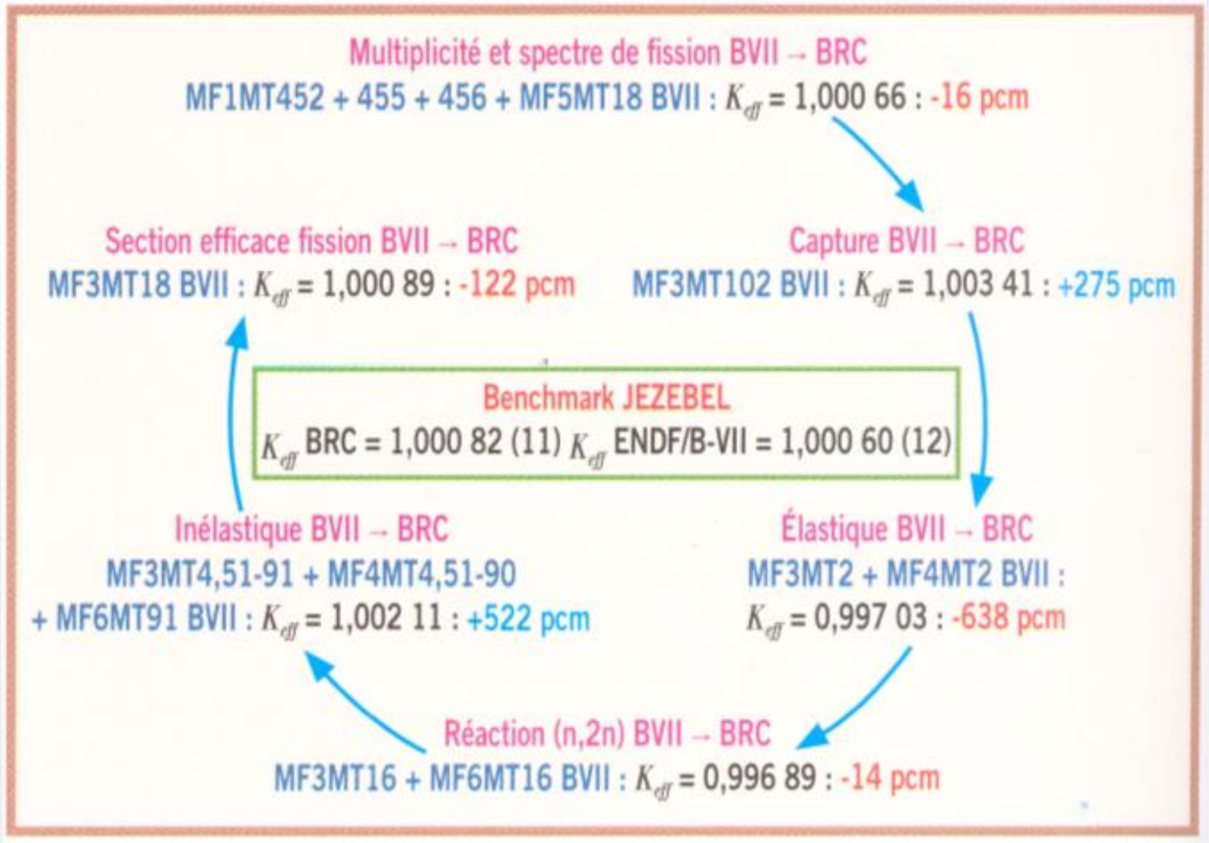
# but you said there would be data??



```
228TH  L  1643.125 15 (2-,3-)                                                    NO HALF-LIFE GIVEN.
228TH  B               0.80  6             7.67   4
228THS B EAV=141.4 10
228TH  G   474.75  10 0.022  3  [M1,E2]                0.15   11
228THS G KC=0.12 9$  LC=0.026 13$  MC=0.006 3$   NC+=0.0023 10
228TH  G   520.151 16 0.067  5  (M1)                   0.201                A    DCC of 3 from theory assumed
228THS G KC=0.161$  LC=0.0301
228TH cG M         |a(K)exp=0.31 {I12} (1971He23); theory: |a(K)=0.161
228TH  G   627.23  20 0.014  3
228TH  G   674.16     0.109  LE [E1]                                        R    +-1 KEV ASSIGNED. RI/2 ASSUMED.
228TH3 G FL=968.968                                                              FL= FOUND.
228TH cG E         deduced from E(level)
228TH  G   674.75     0.109  LE [M1,E2]                0.06   4             R    +-1
228TH3 G FL=968.369                                                              FL=
228THS G KC=0.048 32$  LC=0.010 5
228TH cG E         deduced from E(level)
228TH  G   699.08  15 0.037  5  [D,E2]                 0.05   4
228TH  G  1247.08  4  0.50   3  (M1)                   0.0201               A    DCC
228TH3 G FLAG=IL
228TH  G  1315.34  10 0.015  3  [M1,E2]                0.012  6
228TH  L  1645.954 12  3+                                                        NO H
228TH  B               4.16  20            6.94   2
```



# there are issues with our data…



Multiplicité et spectre de fission BVII → BRC
MF1MT452 + 455 + 456 + MF5MT18 BVII : $K_{eff}$ = 1,000 66 : -16 pcm

Capture BVII → BRC
MF3MT102 BVII : $K_{eff}$ = 1,003 41 : +275 pcm

Élastique BVII → BRC
MF3MT2 + MF4MT2 BVII :
$K_{eff}$ = 0,997 03 : -638 pcm

Réaction (n,2n) BVII → BRC
MF3MT16 + MF6MT16 BVII : $K_{eff}$ = 0,996 89 : -14 pcm

Inélastique BVII → BRC
MF3MT4,51-91 + MF4MT4,51-90
+ MF6MT91 BVII : $K_{eff}$ = 1,002 11 : +522 pcm

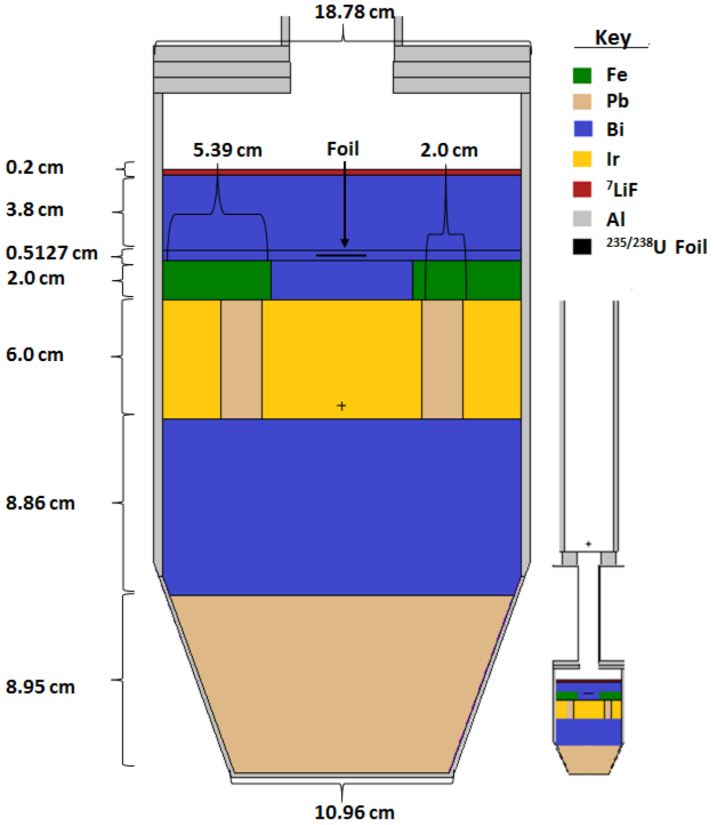
Section efficace fission BVII → BRC
MF3MT18 BVII : $K_{eff}$ = 1,000 89 : -122 pcm

Benchmark JEZEBEL
$K_{eff}$ BRC = 1,000 82 (11) $K_{eff}$ ENDF/B-VII = 1,000 60 (12)

- We tune the knobs on uncertain things to get integral results to work out
- How to deal with improving things?
- How to find the real data?

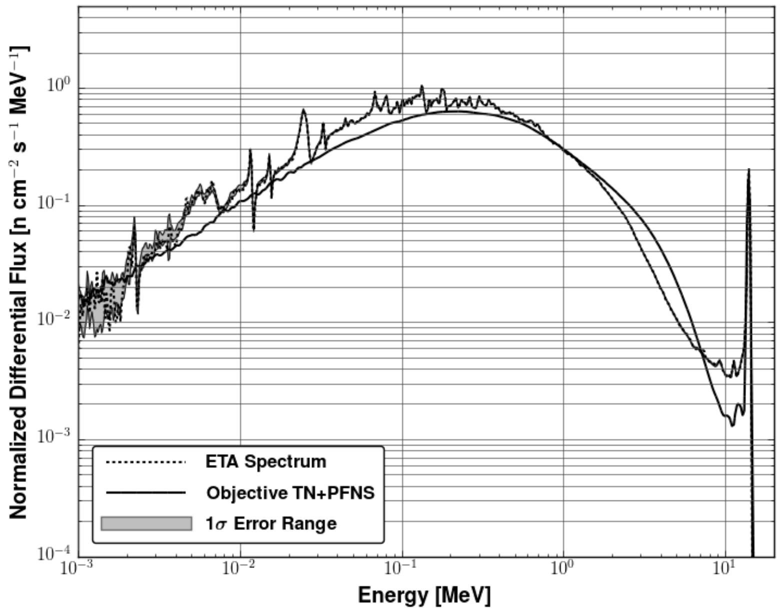
Morgan C. White, “Searching for the ‘Right’ Plutonium.” UC Berkeley Colloquium, LA-UR-14-27739, Berkeley, CA, December 2014.

# project: a tool that will shape spectra

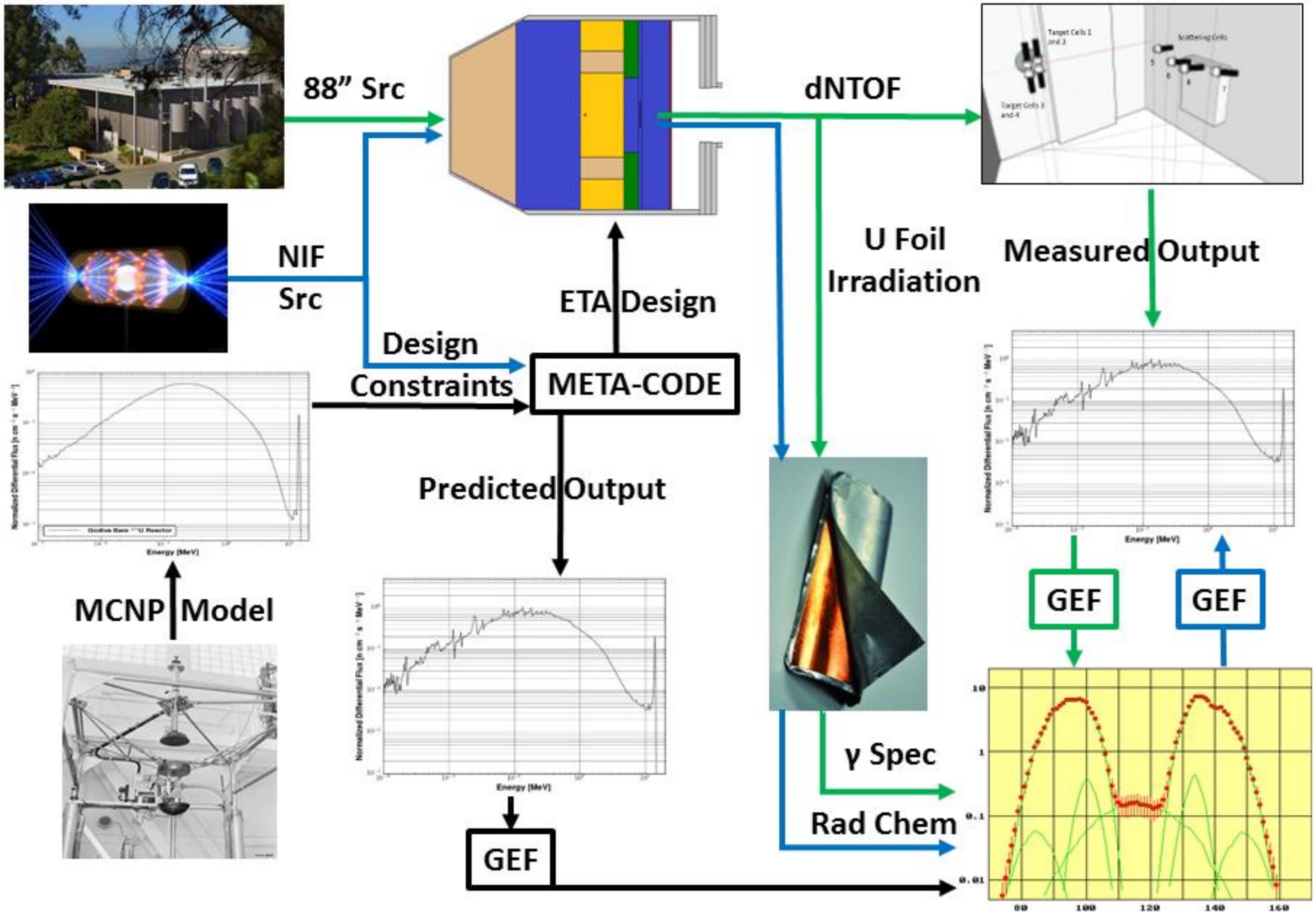


ETA Design

- Use layers of materials to shape neutron energies
- Design by
  - Optimization code
  - Uses predictive computing



# can strategically inform data needs



88” Src

NIF Src

dNTOF

U Foil Irradiation

Measured Output

ETA Design

Design Constraints

META-CODE

Predicted Output

MCNP Model

GEF

GEF

GEF

γ Spec

Rad Chem

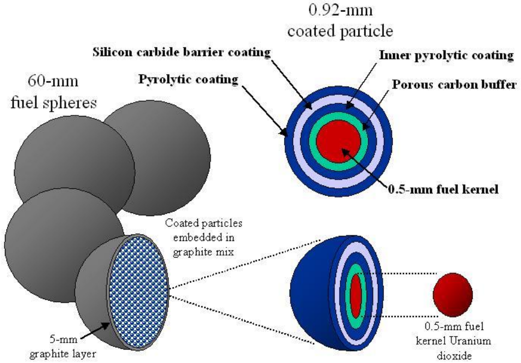
J. Bevins et al., “Application of Metaheuristic Optimization Methods for Neutron Spectral Shaping Applications.” Proceedings of the Conference on Data Analysis 2016 in Santa Fe, NM, March 2016.

# + other important solutions

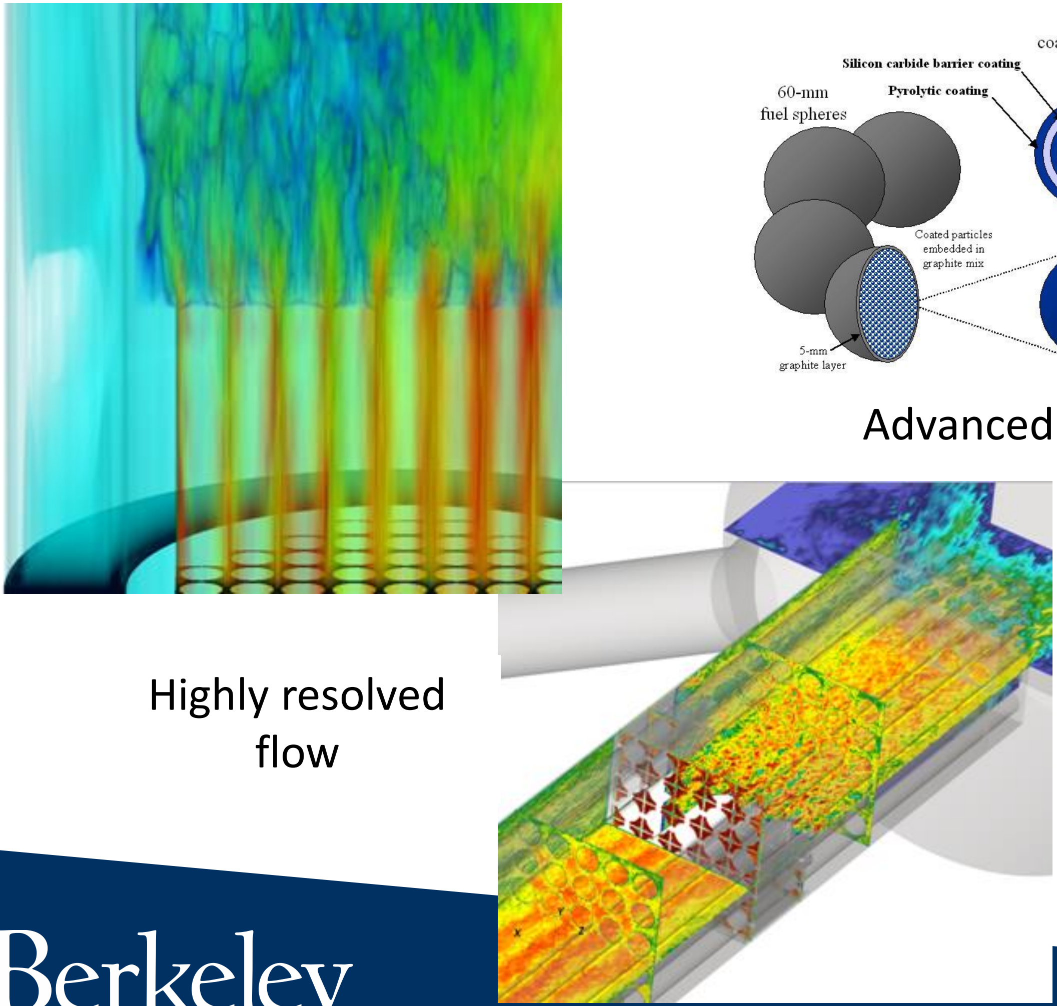
- Community communication
  - People who *make* data and people who *use* data have historically not talked
- Data formats and processing tools that reduce error
  - No one person can actually start from the measured nuclear data and generate what is used by software
  - Some of our data formats are horrible
- Processing the data from experiments that have already been done
  - Lots of data is sitting on a shelf waiting to be analyzed and made available

L. Bernstein, D. Brown, et al. “Nuclear Data Needs and Capabilities for Applications.” White Paper. Lawrence Berkeley National Laboratory, May 27-29 2015.

# soooo much output and analysis



Advanced fuels



Highly resolved flow

Experimental reactors

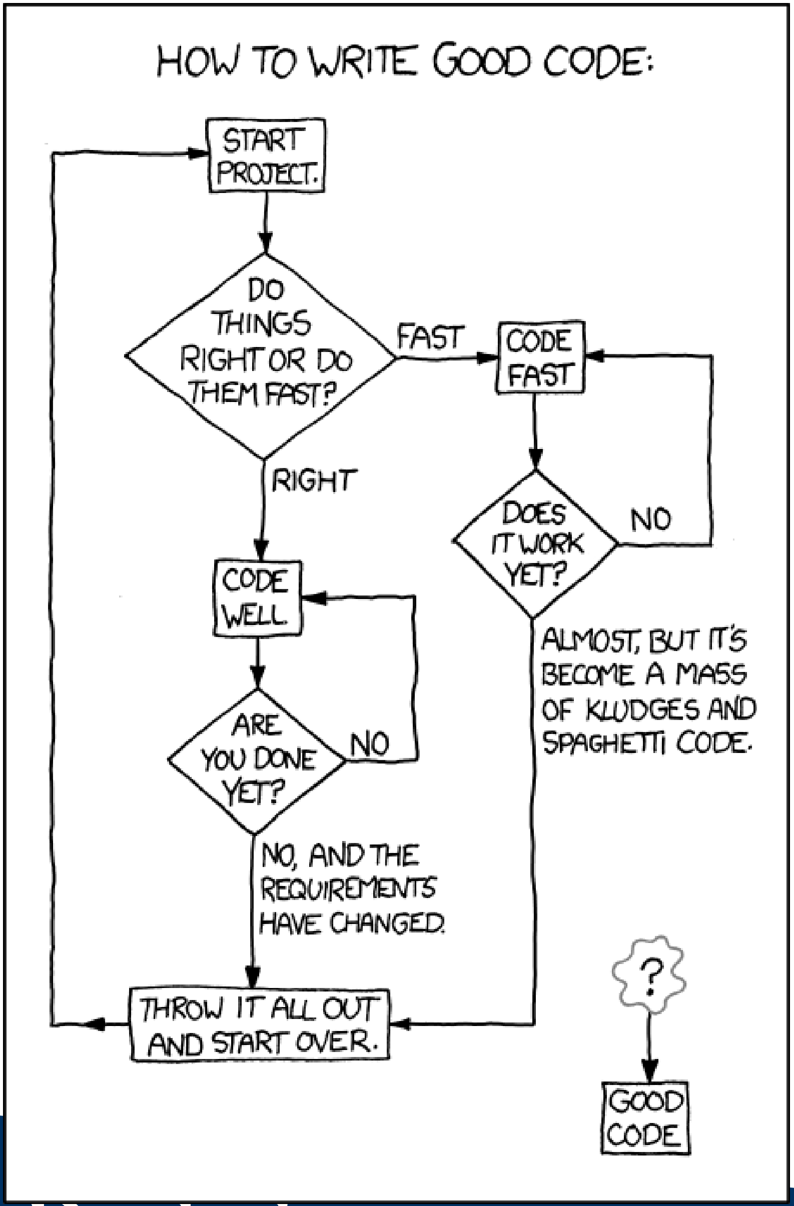
# we need a better way

- Size and scale problem: storing, transferring, visualizing, analyzing, processing, …
  - Getting useful information out of our highly detailed simulations?

- Memory challenges intensified by some new architectures

- What can we do?
  - Write algorithms that process results as they compute?
  - Redesign memory use?
  - Be more strategic in what we ask for and what we report?
  - Time tradeoffs? Output loss?

# P.S. quality software required


HOW TO WRITE GOOD CODE:
START PROJECT.
DO THINGS RIGHT OR DO THEM FAST?
FAST
CODE FAST
DOES IT WORK YET?
NO
ALMOST, BUT IT'S BECOME A MASS OF KLUDGES AND SPAGHETTI CODE.
RIGHT
CODE WELL.
ARE YOU DONE YET?
NO
NO, AND THE REQUIREMENTS HAVE CHANGED.
THROW IT ALL OUT AND START OVER.
?
GOOD CODE

THW

software carpentry

BERKELEY Institute for Data Science

# summary

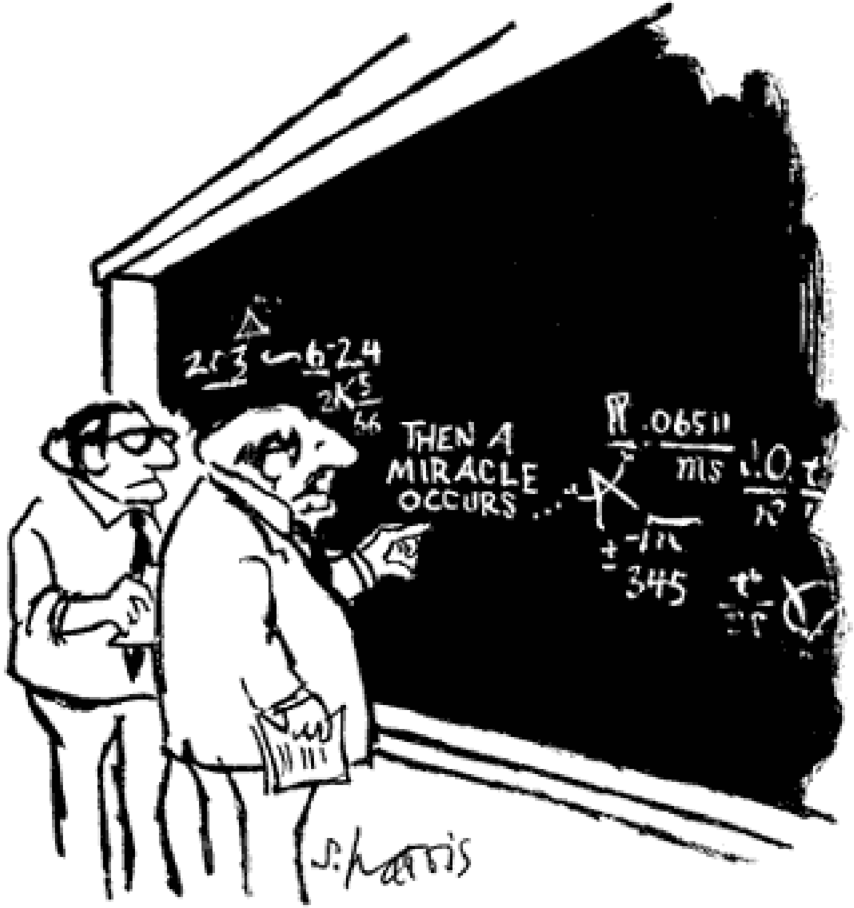
- We need nuclear innovation to help solve some of the world’s biggest challenges
- Predictive simulation is required, but how do we get there?

- Algorithms = physics + hardware
  - Take advantage of large and heterogeneous architectures
  - Combine with physical insights
- Improve nuclear data: experiments, processing tools, processing old data
- Develop new data processing strategies

# acknowledgements

- Ex. 1: Tom Evans, Greg Davidson, Paul Wilson
  Rickover Fellowship, OLCF
- Ex. 2: Kelly Rowland, Ryan Bergmann, Jasmina Vujic
  NSSC, NEUP fellowship
- Ex. 3: Madicken Munk, Tom Evans, Douglas Peplow, Scott Mosher, Tara Pandya, Seth Johnson, Richard Vasques
  Department of Energy award number DE-NE0008286
- Data: James Bevins, Lee Bernstein, Bill Dunlop, Eugene Henry, Dawn Shaughnessy, Morgan White
  NSF Fellowship, LLNL, DTRA, DNDO, NNSA
- Analysis: Kathryn Huff

# questions?



"I think you should be more explicit here in step two."

# block Krylov: Rayleigh Quotient iter.

- Power iteration converges as

$$e^{j+1} = C|{}^{\lambda_2}/_{\lambda_1}|e^j$$

- Where $\sigma(\boldsymbol{A}) \equiv \{\lambda \in \mathbb{C}: \mathrm{rank}(\boldsymbol{A} - \lambda\boldsymbol{I}) \leq n\}$
- We apply a shift, $\mu$, and $(\boldsymbol{A} - \mu\boldsymbol{I})$ has the same eigenvectors as $\boldsymbol{A}$

$$\sigma(\boldsymbol{A} - \mu\boldsymbol{I}) \equiv \{\frac{1}{\lambda - \mu}; \lambda \in \sigma(\boldsymbol{A})\}$$

$$e^{j+1} \approx C|\frac{\lambda_1 - \mu}{\lambda_2 - \mu}|e^j$$

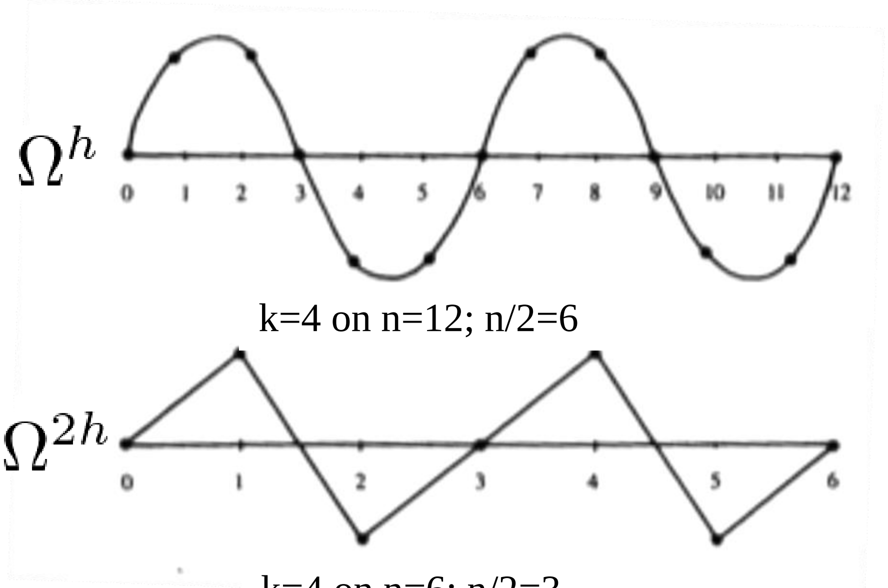
- And make the shift the Rayleigh Quotient (optimal guess)
- *However*, now <u>looks like every group has upscattering</u>!!
- This method only makes sense with Block Krylov

Berkeley UNIVERSITY OF CALIFORNIA

# RQI: multigrid in energy precond.

- Unfortunately, the better the guess the worse the condition #
- RQI (especially when using a Krylov method) requires preconditioning

- Multigrid in energy
- Right preconditioner
- Strategically removes slow-converging error modes
- Easy to implement
- Easy to parallelize in energy



k=4 on n=12; n/2=6

k=4 on n=6; n/2=3