RQI with a Multigrid in Energy Preconditioner for Massively Parallel Neutron Transport

R. N. Slaybaugh[*],
T. M. Evans[+], G. G. Davidson[+], P. P. H. Wilson[†]

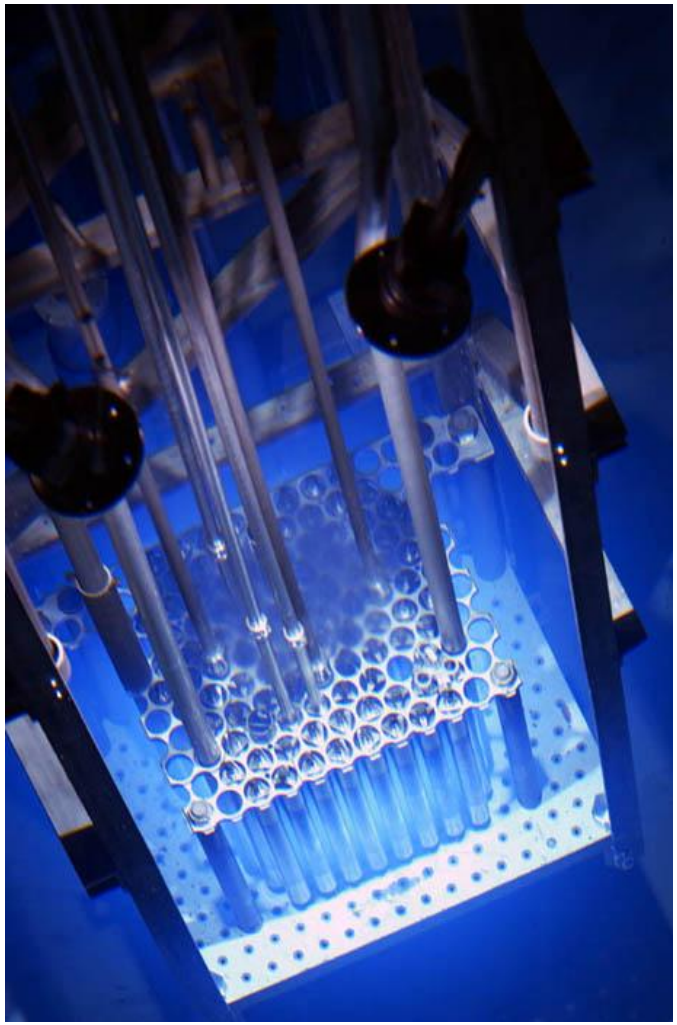
[*]University of California, Berkeley,
[+]Oak Ridge National Laboratory,
[†]University of Wisconsin – Madison

M&C + SNA + MC 2015

Nashville, TN, April 23, 2015

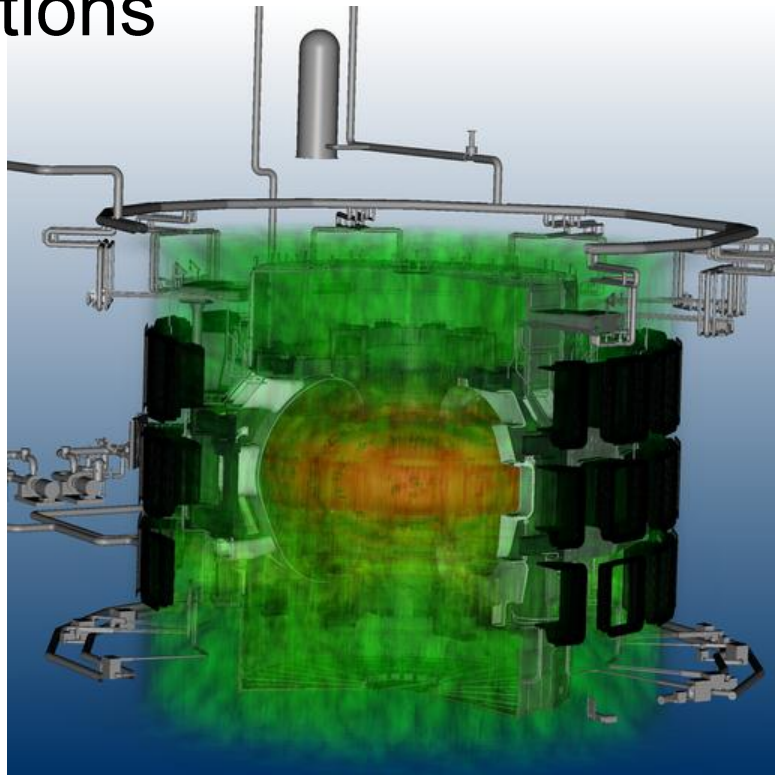
Outline

- Background
- Block Krylov
- RQI
- MGE Preconditioner
- Results
- Conclusions



Large Problems and HPC

- Systems of interest can be geometrically large and/or require fine discretizations
- We also want to do coupled multi-physics
- High performance computers (HPC) have big memories, many cores
- Enable solution of "grand challenge" problems



Major radius of 6.2m

http://www.efda.org/the_iter_project/iter_the_machine.htm

Transport Eqn. Operator Form

- After discretization, operator form of transport eqn:

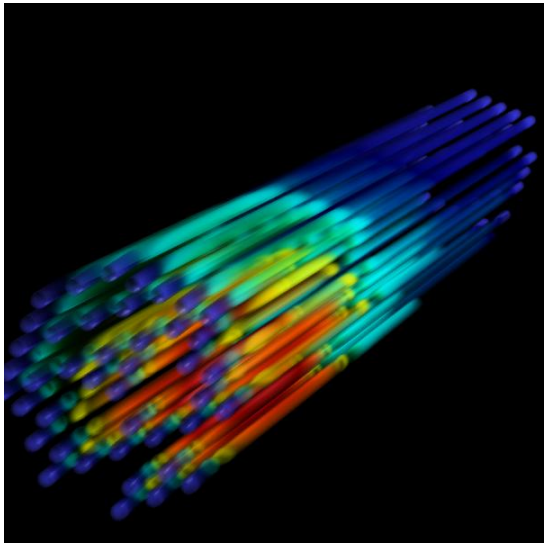
- $\mathbf{L} = \hat{\Omega} \cdot \nabla + \Sigma$ is the transport operator,
- $\mathbf{M}$ converts harmonic moments into discrete angle sets,
- $\mathbf{D} = \mathbf{M}^T \mathbf{W} = \sum_{a=1}^{n} Y_{lm}^{e/o} w_a$
  is the discrete-to-moment operator.
- f contains the fission source, $\nu\Sigma_f$;
- $\mathbf{F} = \chi f^T$,
- $\mathbf{S}$ is the scattering matrix,

$$\mathbf{L}\psi = \mathbf{MS}\phi + \frac{1}{k}\mathbf{MF}\phi$$

$$\phi = \mathbf{D}\psi$$

Denovo's Methods (used here)

- Within-group solvers:
  - Krylov (GMRES, BiCGStab), source iteration
- Multigroup Solvers:
  - Transport Two-Grid upscatter acceleration of Gauss Seidel
  - Block Krylov (GMRES, BiCGStab)
  - Multigrid in energy preconditioning
- Eigenvalue solvers:
  - Power iteration (rebalance, CMFD)
  - Rayleigh Quotient Iteration



Multigroup Solution

- Gauss Seidel does energy iterations over the G space-angle inner iterations
- Space-angle done with Krylov on single group
- Fundamentally serial in energy

- Block Krylov puts a block of groups at the same iteration level
- Handles space-angle and energy all at once
- Krylov should converge more quickly than GS

Block Krylov: One Iteration Level

- Apply **A** to whole upscatter block at once

$$\mathbf{S} = \begin{pmatrix} [\mathbf{S}]_{11} & 0 & 0 & 0 & 0 \\ [\mathbf{S}]_{21} & [\mathbf{S}]_{22} & 0 & 0 & 0 \\ [\mathbf{S}]_{31} & [\mathbf{S}]_{32} & [\mathbf{S}]_{33} & [\mathbf{S}]_{34} & [\mathbf{S}]_{35} \\ [\mathbf{S}]_{41} & [\mathbf{S}]_{42} & [\mathbf{S}]_{43} & [\mathbf{S}]_{44} & [\mathbf{S}]_{45} \\ [\mathbf{S}]_{51} & [\mathbf{S}]_{52} & [\mathbf{S}]_{53} & [\mathbf{S}]_{54} & [\mathbf{S}]_{55} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix}$$

Downscatter Block

Upscatter RHS

Upscatter Block

$$\underbrace{\left(\mathbf{I} - \mathbf{D}\mathbf{L}^{-1}\mathbf{M}\mathbf{S}_{\text{up\_block}}\right)}_{\mathbf{A}}\phi = \underbrace{\mathbf{D}\mathbf{L}^{-1}\mathbf{M}\left(q_e^{\text{up}}\right)}_{b}$$

- Iteration vector includes all upscattering groups
- Enables parallelization in energy (“multisets”)

Power Iteration

- Ordinary eigenvalue equation traditionally solved with Power Iteration (PI) [define $\mathbf{T} = \mathbf{DL}^{-1}$]:

$$\mathbf{A} = (\mathbf{I} - \mathbf{TMS})^{-1}\mathbf{TM}$$

$$\sigma(\mathbf{A}) \equiv \{\lambda \in \mathbb{C} : rank(\mathbf{A} - \lambda\mathbf{I}) \leq n\}$$

$$\mathbf{A}^{j} x_i = \lambda_i^{j} x_i$$

$$\phi^{j+1} = \frac{1}{k^{j}} \mathbf{A}\phi^{j} \qquad k^{j+1} = k^{j} \frac{f^T \phi^{j+1}}{f^T \phi^{j}}$$

- Can be quite slow for dominance ratios close to 1

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

Shifted Inverse Iteration is Faster

- For some shift, μ, $(\mathbf{A} - \mu\mathbf{I})$ has same eigenvectors as **A**
- If invertible, $\sigma([\mathbf{A} - \mu\mathbf{I}]^{-1}) = \{1/(\lambda - \mu) : \lambda \in \sigma(\mathbf{A})\}$
- Eigenvalues near shift become separated

$$\kappa_1 = \frac{1}{\lambda_1 - \mu}\ , \kappa_2 = \frac{1}{\lambda_2 - \mu}\ , \ \cdots\ , \kappa_n = \frac{1}{\lambda_n - \mu}$$

- As $\mu \to \lambda_1$, $\kappa_1 \to \infty$; other terms stay finite

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

- Like power iteration on $(\mathbf{A} - \mu\mathbf{I})^{-1}$

RQI is Faster Still

- For problems that look like **A**x = λ**B**x, the RQ is:

$$\rho(x) = \frac{x^T \mathbf{A} x}{x^T \mathbf{B} x}$$

- **RQI**: shifted inverse iteration with an optimal shift
- Better convergence properties than PI

$$(\mathbf{I} - \mathbf{T}\mathbf{M}\tilde{\mathbf{S}})\phi = (\gamma - \rho)\mathbf{T}\mathbf{M}\mathbf{F}\phi$$

$$\tilde{\mathbf{S}} \equiv \mathbf{S} + \rho\mathbf{F}$$

- Shifted matrix looks like an energy-block dense scattering

Block Krylov and RQI

- Shifted system would be difficult for GS
- Systems that are block-dense in energy are still sparse in energy-space-angle
- Structurally ideal for Krylov methods
- However, RQI creates *poorly conditioned* systems:

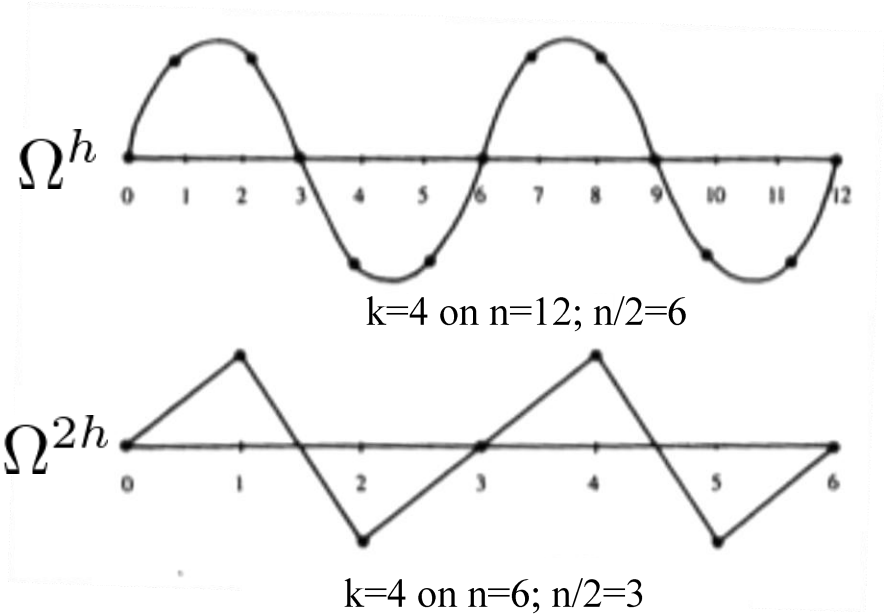
$$\kappa(\mathbf{A}) = ||\mathbf{A}||\ ||\mathbf{A}^{-1}||$$

- Krylov methods can have difficulty
- Result: cannot get a valid RQ; *RQI doesn't converge*

Making Error Oscillatory

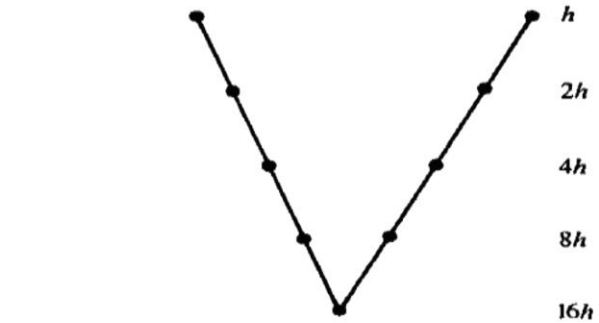
- Key to multigrid: <u>smooth modes look oscillatory on coarser grids</u>; remove them

- Map error from fine grid to coarse grid
- Smooth error is now relatively oscillatory
- Relax on coarse grid to remove this error



- Map solution back to fine grid and correct

Multigrid in Energy



V-Cycle

W-Cycle

Full MG

- Multigrid in Energy (MGE) as right preconditioner
- Relaxation method is weighted Richardson

$$x^{i+1} = \omega \mathbf{TMS} x^i + \omega b + (1-\omega)x^i$$

  - Relaxations per level
  - Number of levels
  - omega

Parallelization of MGE

- With one set, number of grid levels is

$$\text{floor}\left(\log_2(G-1)\right)+2$$

- Multisets: each set does its own “mini” V-cycle
- No communication beyond upscattering required
- Grid depth becomes

$$num\_g_{min} = \text{floor}\left(\frac{num\_groups}{num\_sets}\right)$$

$$num\_grids = \text{floor}\left(\log_2(num\_g_{min}-1)\right)+2$$

Highlights From the Past

- RQI doesn't really work without preconditioning
- Parameter selection: start with w1r2v2
- Use reduced angle set in preconditioner
- Shallow grid depth is sufficient
- Preconditioner scales very well in energy
- Preconditioned PI not faster than PI alone for

- R. N. Slaybaugh, T. M. Evans, G. G. Davidson, and P. P. H. Wilson, "Rayleigh Quotient Iteration in 3D, Deterministic Neutron Transport," *PHYSOR 2012 Advances in Reactor Physics Linking Research, Industry, and Education,* Knoxville, TN, 2012, American Nuclear Society.
- R. N. Slaybaugh, T. M. Evans, G. G. Davidson, and P. P. H. Wilson, "Multigrid in energy preconditioner for Krylov solvers," *Journal of Computational Physics,* **242**, pp. 405–419 (2013).

Questions

1. Will preconditioning with MGE facilitate the use of RQI?

2. Will the combination of RQI, MGE, and the block Krylov solver be advantageous for at least some problems?

MGE Does Not Help PI

- 2D C5G7 Benchmark
- 16 cores, tolerance 10^{-3}, k tolerance 10^{-5}

Solver	Precond	Krylov	Eigen	Time(s)
PI	none	3,129	32	8.54×10^3
PI	w1.4r2v2	438	31	1.77×10^4
PI	w1r3v3	253	31	2.28×10^4
RQI	none	n/a	n/a	n/a
RQI	w1r3v3	299	19	2.57×10^4

MGE Helps RQI

- 2D C5G7 Benchmark
- 16 cores, tolerance 10^{-3}, k tolerance 10^{-5}

Solver	Precond	Krylov	Eigen	Time(s)
PI	none	3,129	32	8.54×10^{3}
PI	w1.4r2v2	438	31	1.77×10^{4}
PI	w1r3v3	253	31	2.28×10^{4}
RQI	none	n/a	n/a	n/a
RQI	w1r3v3	299	19	2.57×10^{4}

But PI Is Still Better

- 2D C5G7 Benchmark
- 16 cores, tolerance 10^{-3}, k tolerance 10^{-5}

Solver	Precond	Krylov	Eigen	Time(s)
PI	none	3,129	32	8.54×10^3
PI	w1.4r2v2	438	31	1.77×10^4
PI	w1r3v3	253	31	2.28×10^4
RQI	none	n/a	n/a	n/a
RQI	w1r3v3	299	19	2.57×10^4

Problems Are Too Easy; Use PI

- 3D C5G7 Benchmark, DR 0.7709
- 720 cores, tolerance 10^{-4}, k tolerance 10^{-5}

Solver	Precond	Krylov	Eigen	Time(s)
PI	none	3,129	32	4.46×10^3
RQI	w1.3r2v2	302	19	2.32×10^4
PI	w1.3r2v2	288	32	2.84×10^4
RQI*	w1r3v3	103	9	3.02×10^4
PI*	w1r3v3	126	14	4.04×10^4
RQI	w1.5r3v3	187	19	3.24×10^4
PI	w1.5r3v3	192	32	3.73×10^4

*different tolerances

Full PWR-900 Details

2 x 2 spatial cells/pin

17 x 17 pins/assembly

289 assemblies (132 reflector, 159 fuel of varying enrichment)

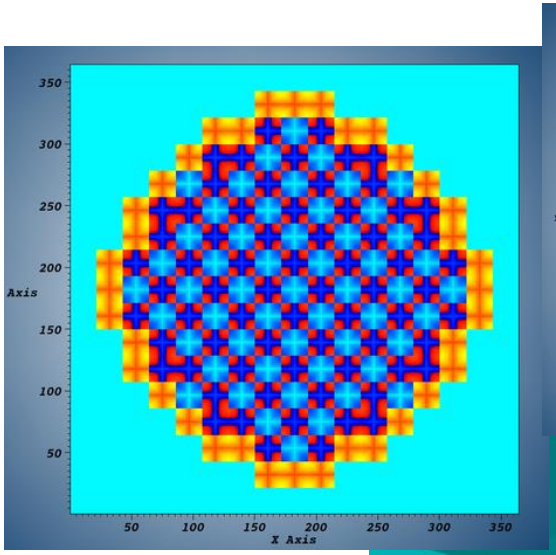
P_0: 1 moment

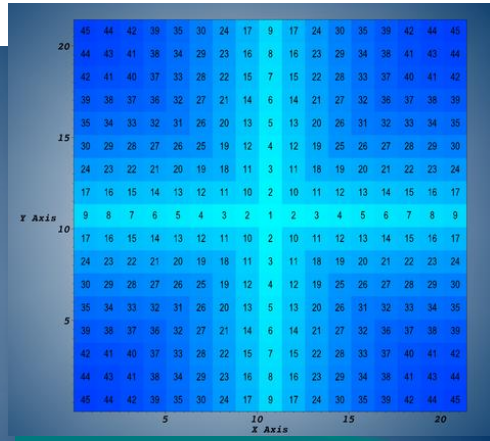
S_{12}: 168 angle sets (MGE: S_2)

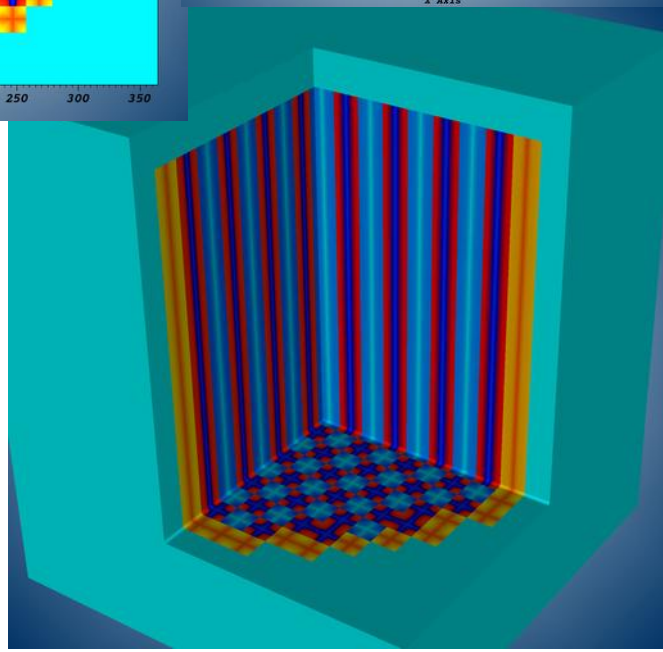
233,858,800 cells; 1 unknown/cell

44 groups: *1.73 trillion unknowns*

12,544 blocks; 137,984 cores (11 sets)







RQI Can Beat PI

- RQI needed **less time** and **fewer iterations**

Solver	Precond	Krylov	Eigen	Time (m)
PI	none	5.602	149	612.2
PI	w1r2v2	946	86	720*
PI	w1r3v3	111	11	480*,+
RQI	w1r2v2	70	5	54.8
RQI	w1r3v3	76	6	330.4+

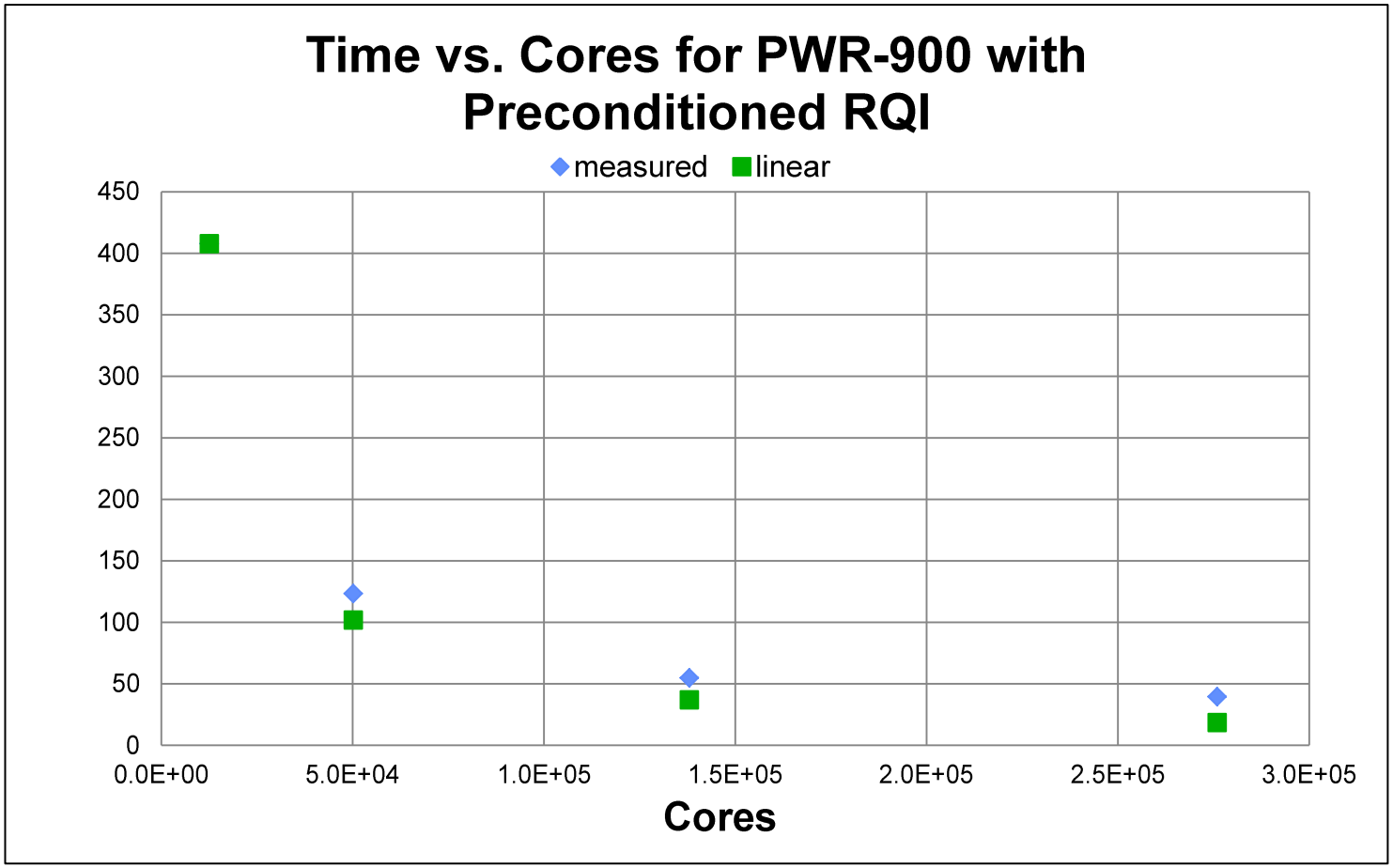
*Exceeded wall time limit

+S_{12} in MGE; different tolerances and decomposition

RQI+MGE Strong Scaling

Sets	Domains	Time (m)	Tperfect	Efficiency
1	12,544	407.8	407.8	1.000
4	50,176	123.4	102.0	0.826
11	137,984	54.8	37.1	0.676
22	275,968	39.6	18.5	0.468

RQI+MGE Strong Scaling

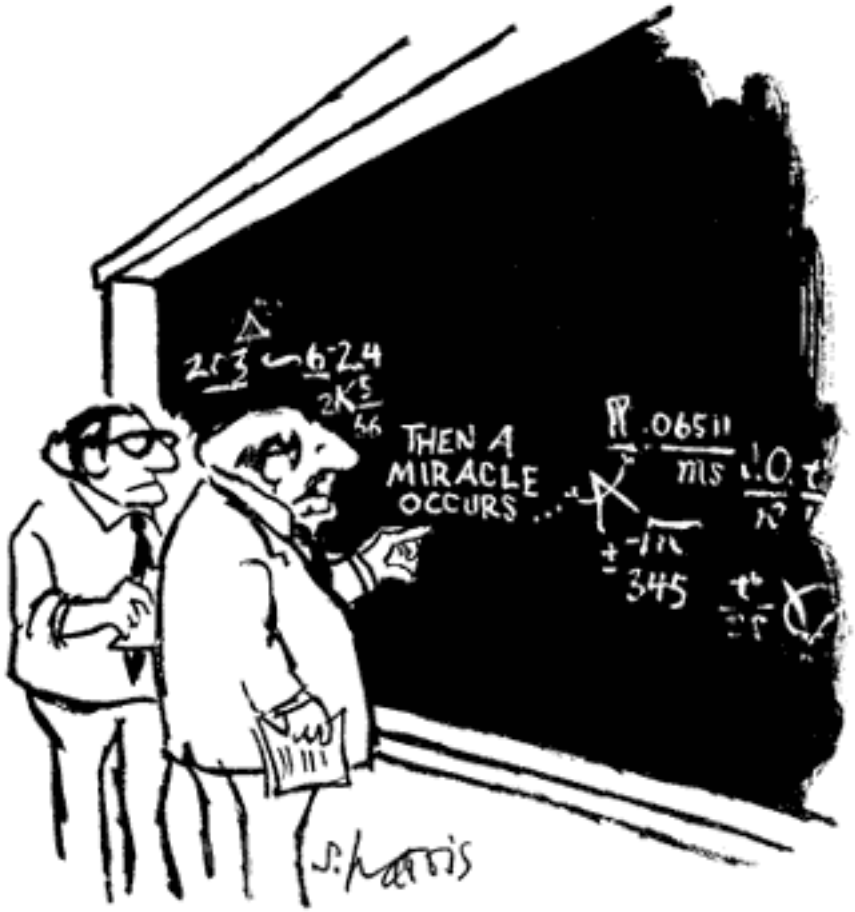


Conclusions

- RQI is an optimal eigenvalue method
  - Desirable for loosely coupled systems
  - Needs preconditioning for large problems
  - Can be a win for challenging problems

- Is enabled by multigroup Krylov solver
- Multigrid in Energy preconditioner can converge flux so RQI can converge challenging problems
- The methods used in concert performed well on large machines for real problems

Questions?



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

Parameter Effects?

- User input parameters:
  - w = Richardson relaxation weight
  - r = number of relaxations per grid
  - v = number of concatenated V-cycles
- w = 1 always safe; up to 1.3 or 1.4 often beneficial
- Increasing r and v always reduced Krylov count
- r1v2 vs. r2v1
  - same impact on iteration count
  - may result in different times

Small Test Success 1

- 3 x 3 x 3, 0.1 grid size, vacuum BCs, 2 materials, 4 downscattering-only groups, S_2, P_0

  k_{ref} = 0.11752, dominance ratio = 1.396e-1

	PI	RQI
k	0.11752	0.11752
group iters	2 or 3 / group	4 to 8
eigen iters	7	6
total Krylov	73	39

Small Test Success 2

- 3 x 3 x 3, 0.1 grid size, reflecting BCs, 1 material, 4 downscattering-only groups, S_2, P_0

  k_{ref} = 2, dominance ratio = 1.630e-15

	PI	RQI
k	2	2
group iters	9 / group	17, 18
eigen iters	2	2
total Krylov	72	35

Transport Equation

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

- $\psi(\vec{r}, \hat{\Omega}, E)$ is the angular neutron flux in neutrons per unit length squared per steradian,
- $\chi(E)$ is the fission spectrum,
- k is the eigenvalue, which can be thought of as the asymptotic ratio of the number of neutrons in one generation to the number in the next,
- Σs are probabilities of interaction with units of inverse length,
- ν is the average number of neutrons released per fission.

Krylov Methods are Powerful

- Solutions to ***A****x* = *b* formed iteratively from Krylov subspace:

$$\mathcal{K}_k(\mathbf{A}, v_1) \equiv span\{v_1, \mathbf{A}v_1, \mathbf{A}^2v_1, ..., \mathbf{A}^{k-1}v_1\}$$

Useful:

  - because robust, often converge quickly, easy to precondition, only matrix-vector products
  - when **A** is large, not explicitly formed, sparse
- Drawbacks: can create large subspaces
- Restart methods available

Gauss Seidel for Outer Iterations

- Scattering to own group, lower group, higher group
- GS does the outer (energy) iteration:
- Inner (space-angle) iteration for 1 to G:

$$\underbrace{\left(\mathbf{I}-\mathbf{DL}^{-1}[\mathbf{M}][\mathbf{S}]_{gg}\right)}_{\text{A}}\underbrace{[\phi]_g^*}_{\text{x}}=\underbrace{\mathbf{DL}^{-1}[\mathbf{M}]\left(\sum_{g'=1}^{g-1}[\mathbf{S}]_{gg'}[\phi]_{g'}^{new}+\sum_{g'=g+1}^{G}[\mathbf{S}]_{gg'}[\phi]_{g'}^{old}+[q_e]_g\right)}_{\text{b}}$$

- Update right hand side, repeat loop for upscattering groups until convergence

Krylov for Inner Iterations

- The Krylov solver applies ***A*** to a group-sized iteration vector, *v*, which represents ϕ :

The action of $\mathbf{A}$ is implemented by doing the following for a group g:

$$\underbrace{\left(\mathbf{I} - \mathbf{D}\mathbf{L}^{-1}[\mathbf{M}][\mathbf{S}]_{gg}\right)}_{\mathbf{A}}[\phi]_g^* =$$

1. matrix-vector multiply: $y_g = [\mathbf{M}][\mathbf{S}]_{gg} v_g,$
2. sweep: $z_g = \mathbf{D}\mathbf{L}^{-1} y_g,$
3. return: $v_g \leftarrow v_g - z_g.$

Block Krylov: One Iteration Level

- Apply ***A*** to whole upscatter block at once

$$\mathbf{S} = \begin{pmatrix} [\mathbf{S}]_{11} & 0 & 0 & 0 & 0 \\ [\mathbf{S}]_{21} & [\mathbf{S}]_{22} & 0 & 0 & 0 \\ [\mathbf{S}]_{31} & [\mathbf{S}]_{32} & [\mathbf{S}]_{33} & [\mathbf{S}]_{34} & [\mathbf{S}]_{35} \\ [\mathbf{S}]_{41} & [\mathbf{S}]_{42} & [\mathbf{S}]_{43} & [\mathbf{S}]_{44} & [\mathbf{S}]_{45} \\ [\mathbf{S}]_{51} & [\mathbf{S}]_{52} & [\mathbf{S}]_{53} & [\mathbf{S}]_{54} & [\mathbf{S}]_{55} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix}$$

Downscatter Block

Upscatter RHS

Upscatter Block

$$\underbrace{\left(\mathbf{I} - \mathbf{D}\mathbf{L}^{-1}\mathbf{M}\mathbf{S}_{\text{up\_block}}\right)}_{\mathbf{A}} \phi^* = \underbrace{\mathbf{D}\mathbf{L}^{-1}\mathbf{M}\left(q_e^{\text{up}}\right)}_{b}$$

- Iteration vector includes all upscattering groups

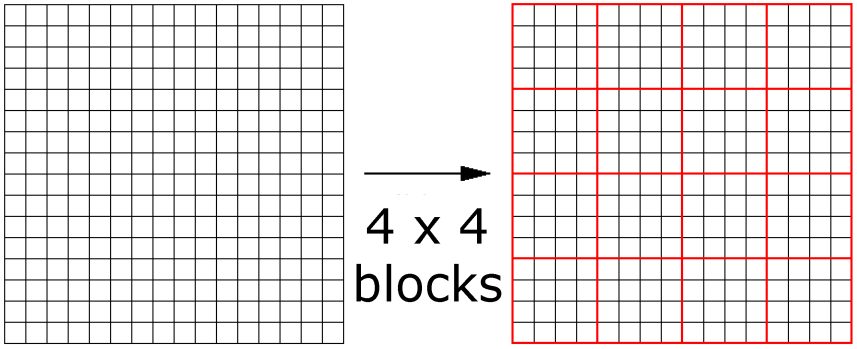
Enables Parallelization in Energy

- Matrix-vector multiply for each upscatter group can be done in parallel
- Communicate result after each multiply
- If each upscatter group computed simultaneously:

$$s_g = \mathbf{S}_{g1} v_1 + \mathbf{S}_{g2} v_{g2} + \cdots + \mathbf{S}_{g5} v_5$$

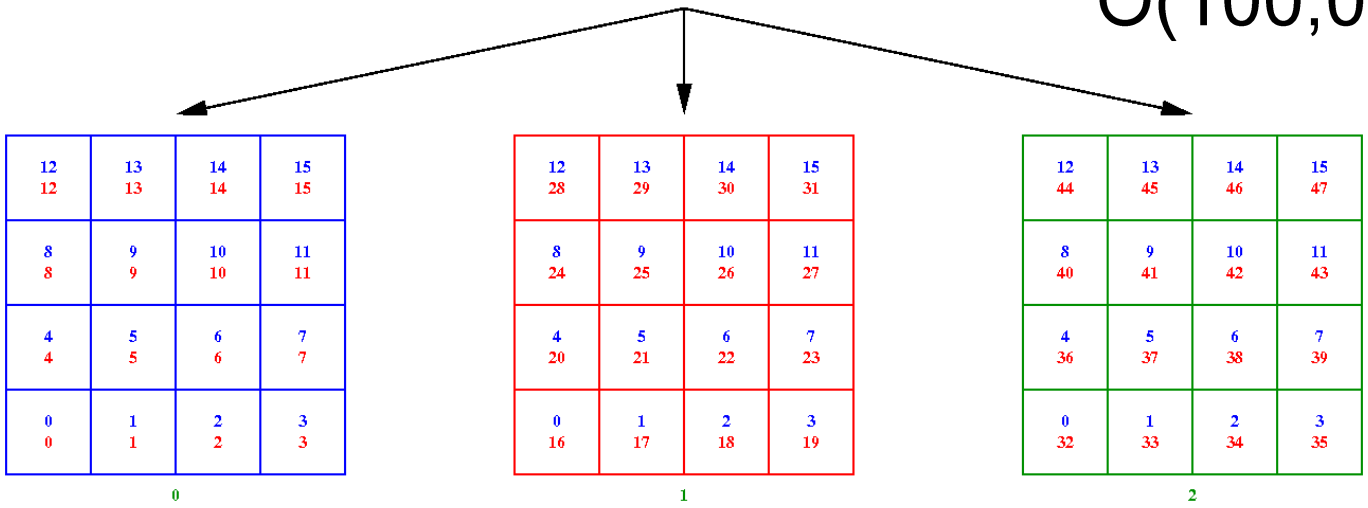
$$\begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{pmatrix} = \begin{pmatrix} [\mathbf{S}]_{11} & 0 & 0 & 0 & 0 \\ [\mathbf{S}]_{21} & [\mathbf{S}]_{22} & 0 & 0 & 0 \\ [\mathbf{S}]_{31} & [\mathbf{S}]_{32} & [\mathbf{S}]_{33} & [\mathbf{S}]_{34} & [\mathbf{S}]_{35} \\ [\mathbf{S}]_{41} & [\mathbf{S}]_{42} & [\mathbf{S}]_{43} & [\mathbf{S}]_{44} & [\mathbf{S}]_{45} \\ [\mathbf{S}]_{51} & [\mathbf{S}]_{52} & [\mathbf{S}]_{53} & [\mathbf{S}]_{54} & [\mathbf{S}]_{55} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix}$$

Energy Set Decomposition



- cores = domains = **energy sets x spatial blocks**
- Code can scale to O(100,000) cores

3 sets



domains = 16 x 3

0 Block id
0 Domain id
0 Set id

Large Test Failure

- 2-D C5G7 MOX benchmark problem
- External mesh file, 2 reflecting BCs, 10 materials, 7 groups w/ 4 upscattering, S_2, P_0

k_{ref} = 1.18655 ± 0.008, dominance ratio = 0.7709

	PI + TTG GS	PI + MG Krylov	RQI
k	1.18538	1.18702	~0.95
mg iters	157 GS	~100	1000
eigen iters	32	32	120*
total Krylov	21,365	3,129	119,006

*terminated manually

RQI Convergence Issues

- The multigroup iterations do not converge