Lawrence Livermore National Laboratory

Chemical Kinetics Research on HCCI & Diesel Fuels and Computationally Efficient Modeling of High-Efficiency Clean Combustion Engines

Dan Flowers, Bill Pitz, Marco Mehl, Mani Sarathy, Charlie Westbrook, Salvador Aceves, Nick Killingsworth, Matt McNenly, Tom Piggott, Mark Havstad, Russell Whitesides

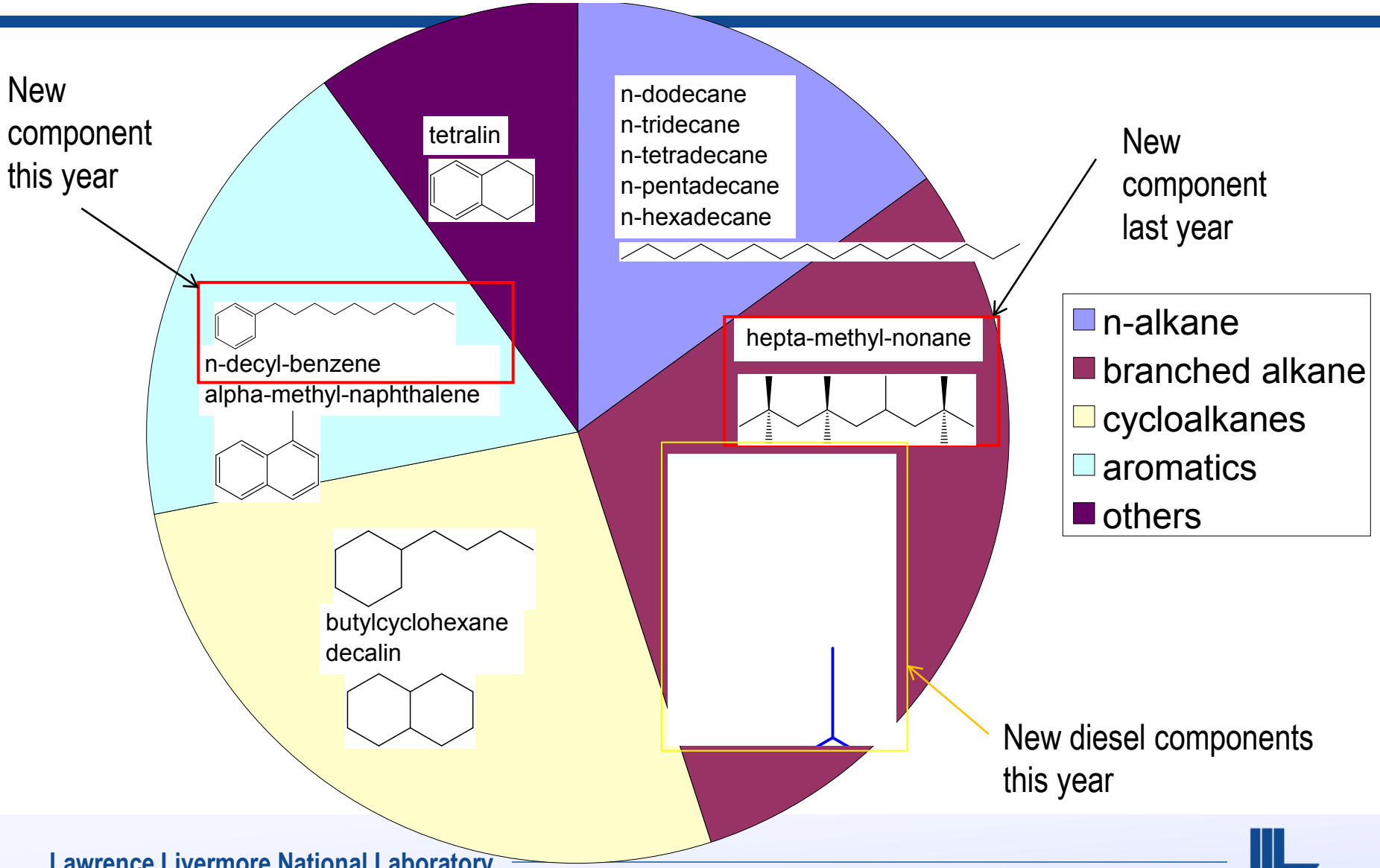
DEER Conference

September 27, 2010 – Detroit, MI

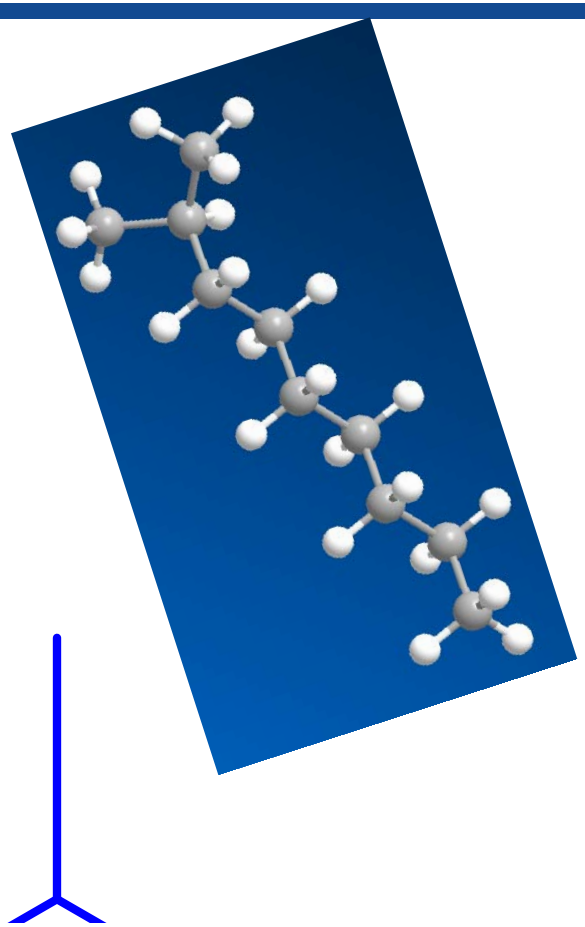
Sponsor: VTP – Team Leaders Gurpreet Singh and Kevin Stork

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344

We have developed mechanisms for complex long-chain species, enabling more representative diesel surrogates



We have developed 2-methyl alkane mechanisms up to C20; branched iso-alkanes are significant components in gasoline and diesel fuels



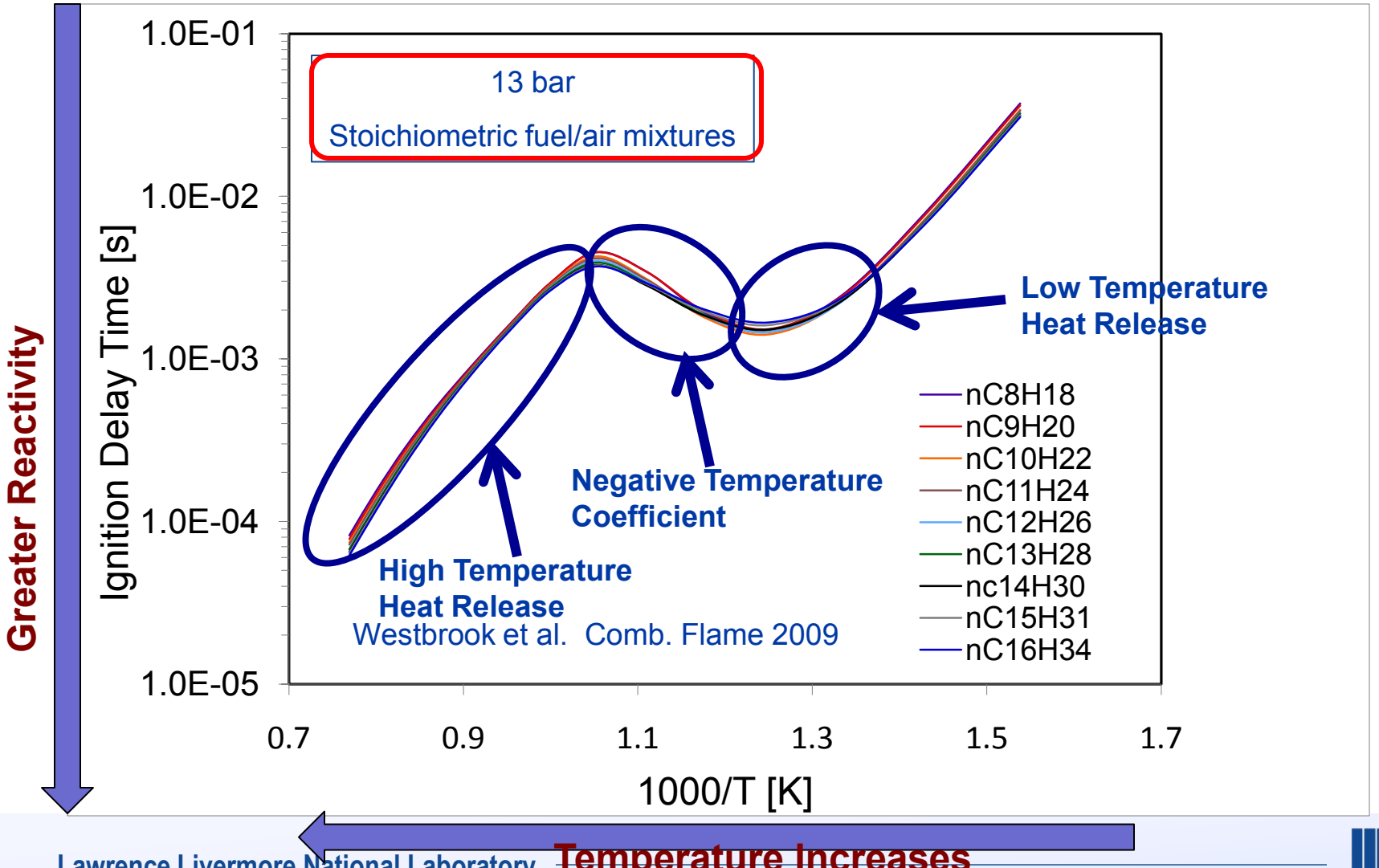
Includes all 2-methyl alkanes up to C20 which covers the entire distillation range for gasoline and diesel fuels

Built with the same reaction rate rules as our successful iso-octane and iso-cetane mechanisms.

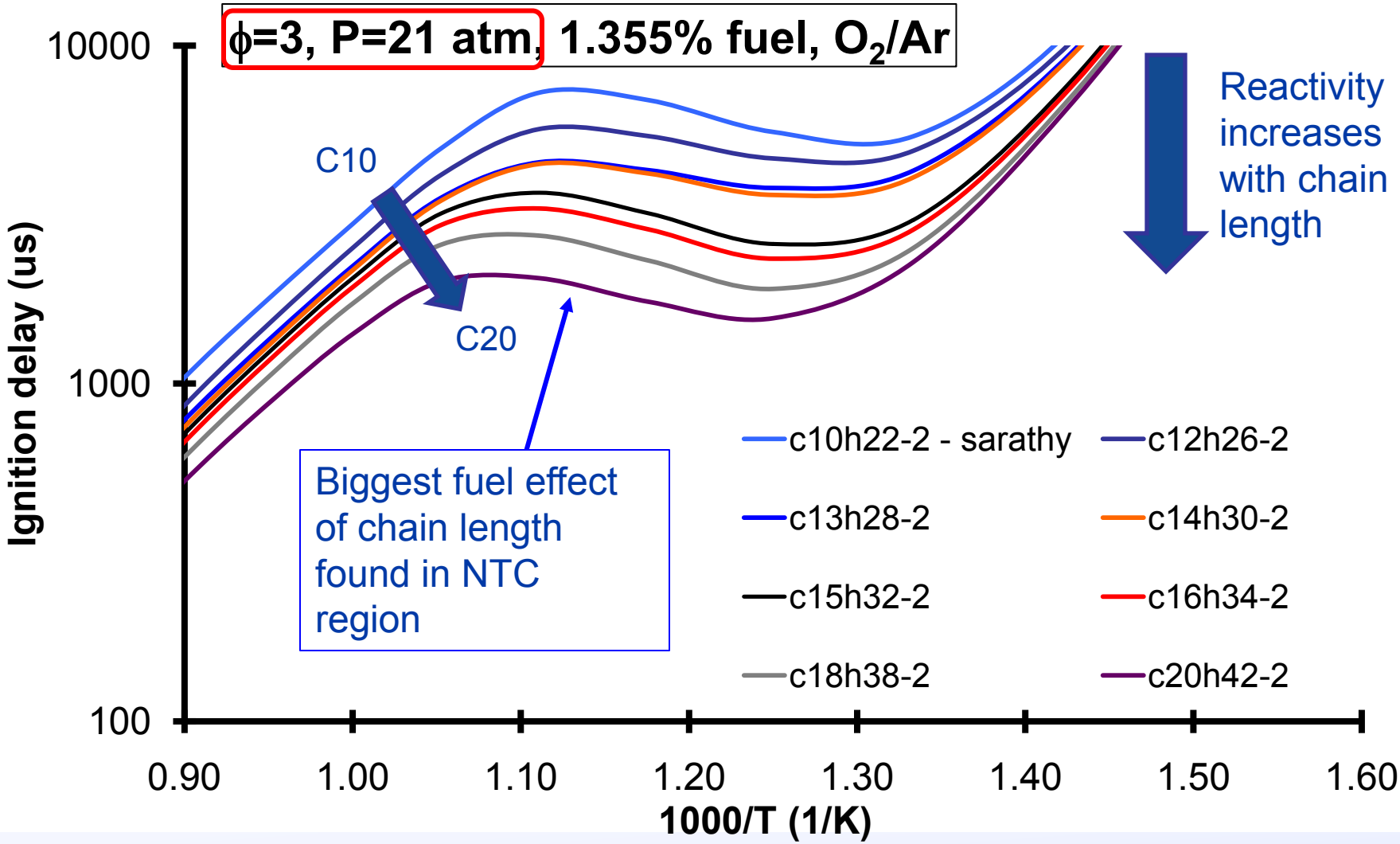
7,900 species

27,000 reactions

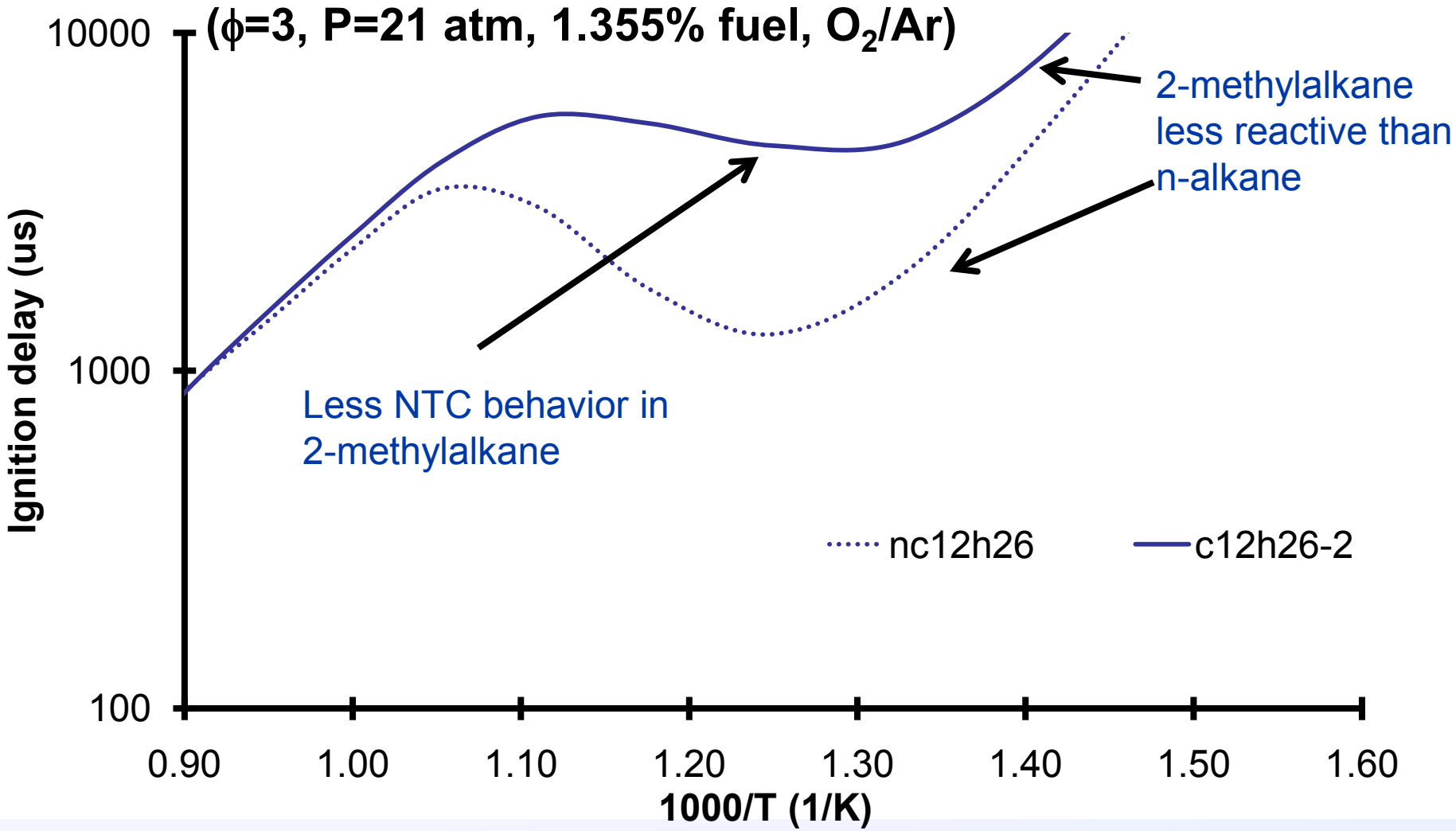
Our previous work showed that ignition characteristics of normal alkanes depend little on carbon chain length



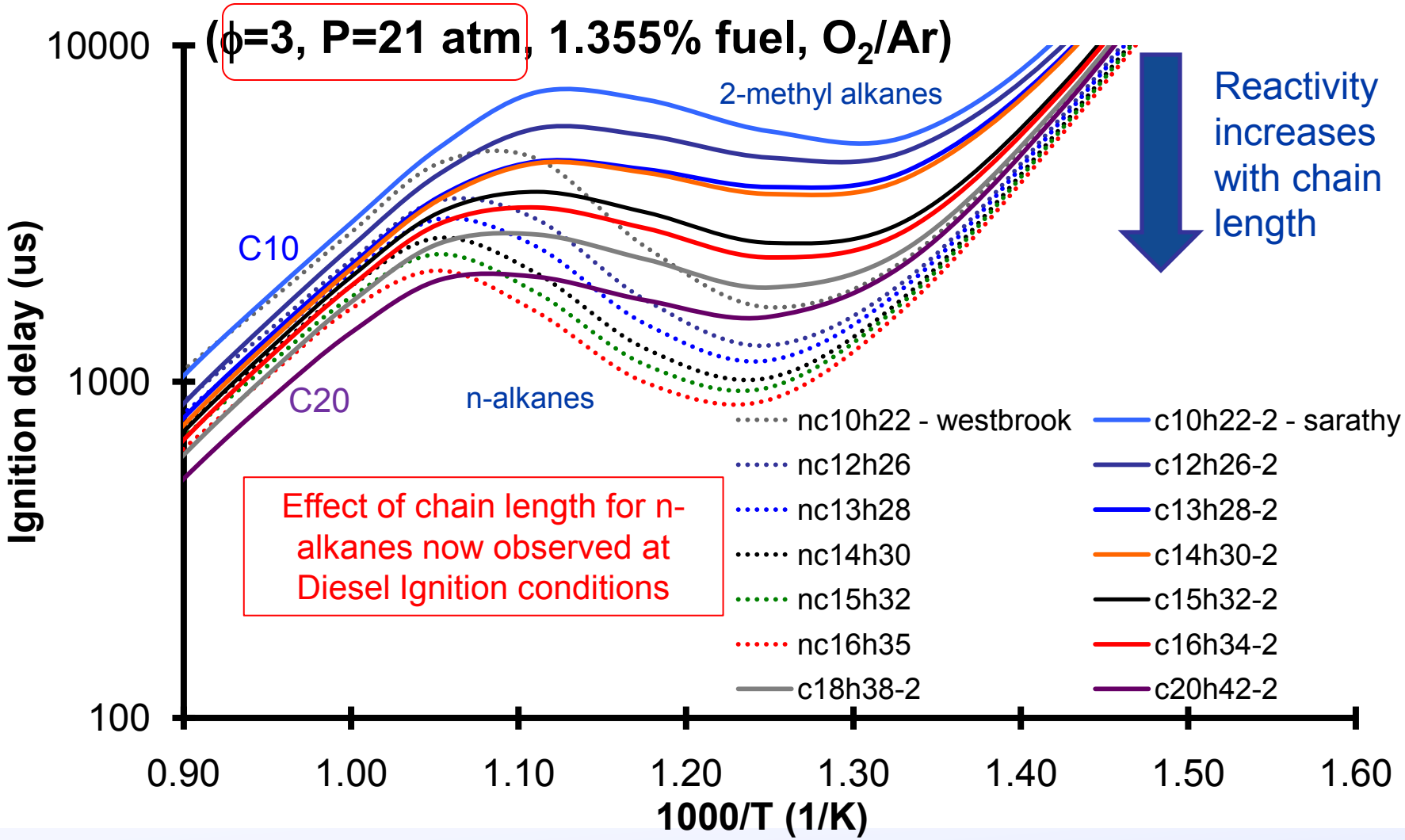
However, 2-methyl alkanes show more reactivity with chain length at higher equivalence ratio and pressure:



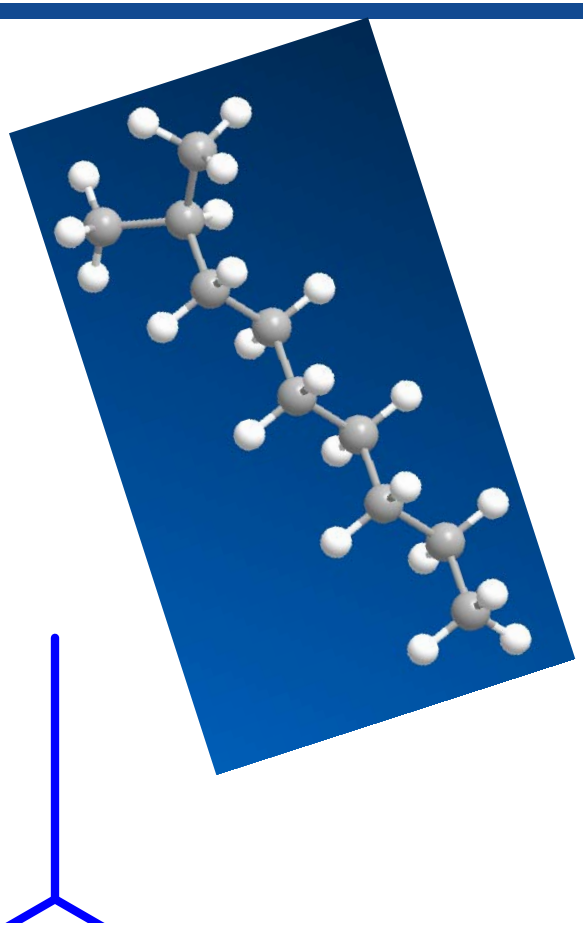
In general, 2-methylalkanes show less negative temperature coefficient (NTC) behavior than n-alkanes



At rich conditions, n-alkanes show chain-length reactivity sensitivity



As mechanisms continue to grow in sophistication, the application of these mechanisms becomes more computationally intensive



Includes all 2-methyl alkanes up to C20 which covers the entire distillation range for gasoline and diesel fuels

Built with the same reaction rate rules as our successful iso-octane and iso-cetane mechanisms.

7,900 species

27,000 reactions

With full (non-sparse) solvers, numerical cost scales with the (# of Species)^3

Solving detailed chemical kinetics combined with 3D fluid dynamics is computationally (and financially) expensive!

Current state-of-the-art mechanism

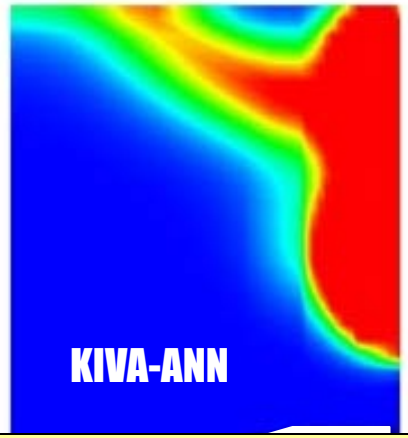
- 7900 species needed for the elementary reaction mechanism for 2-methyl alkanes (Sarathy *et al*. 2010)
- 2M-10M fluid cells for full 3D cylinder model (2010, CERFACS/IFP)

Fully-coupled computation cost: 42,000 Pflop

- If the fastest computer in the world was ideally utilized:
  - **30 hrs on Jaguar (ORNL)** – 240,000 core 1.76 Pflop/s (Number 1 on Top500 supercomputer list)
- **60 years on a workstation** – six core AMD Opteron 100 Gflop/s

Combustion chemistry computational cost is the biggest barrier to complete physics-based simulation suitable for engine design.

We seek to bring the most physically accurate combustion models for the lowest cost

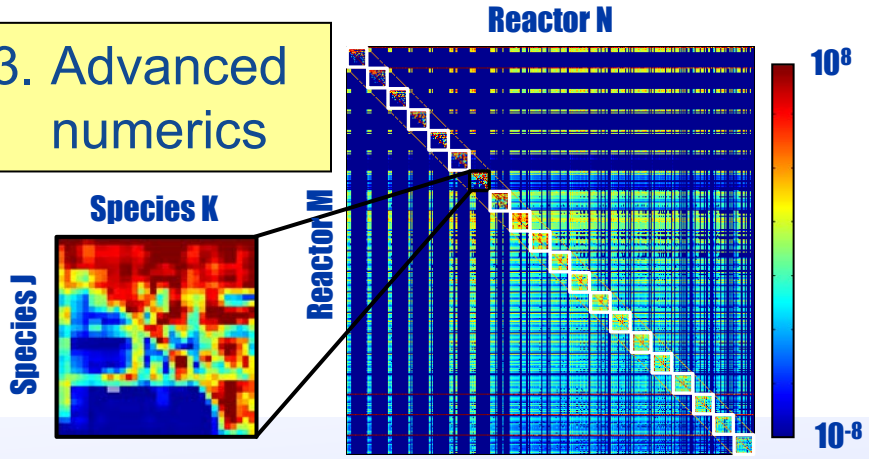


1. Lower-cost models:
   - Reduced Mechanisms
   - ANN Ignition Integral

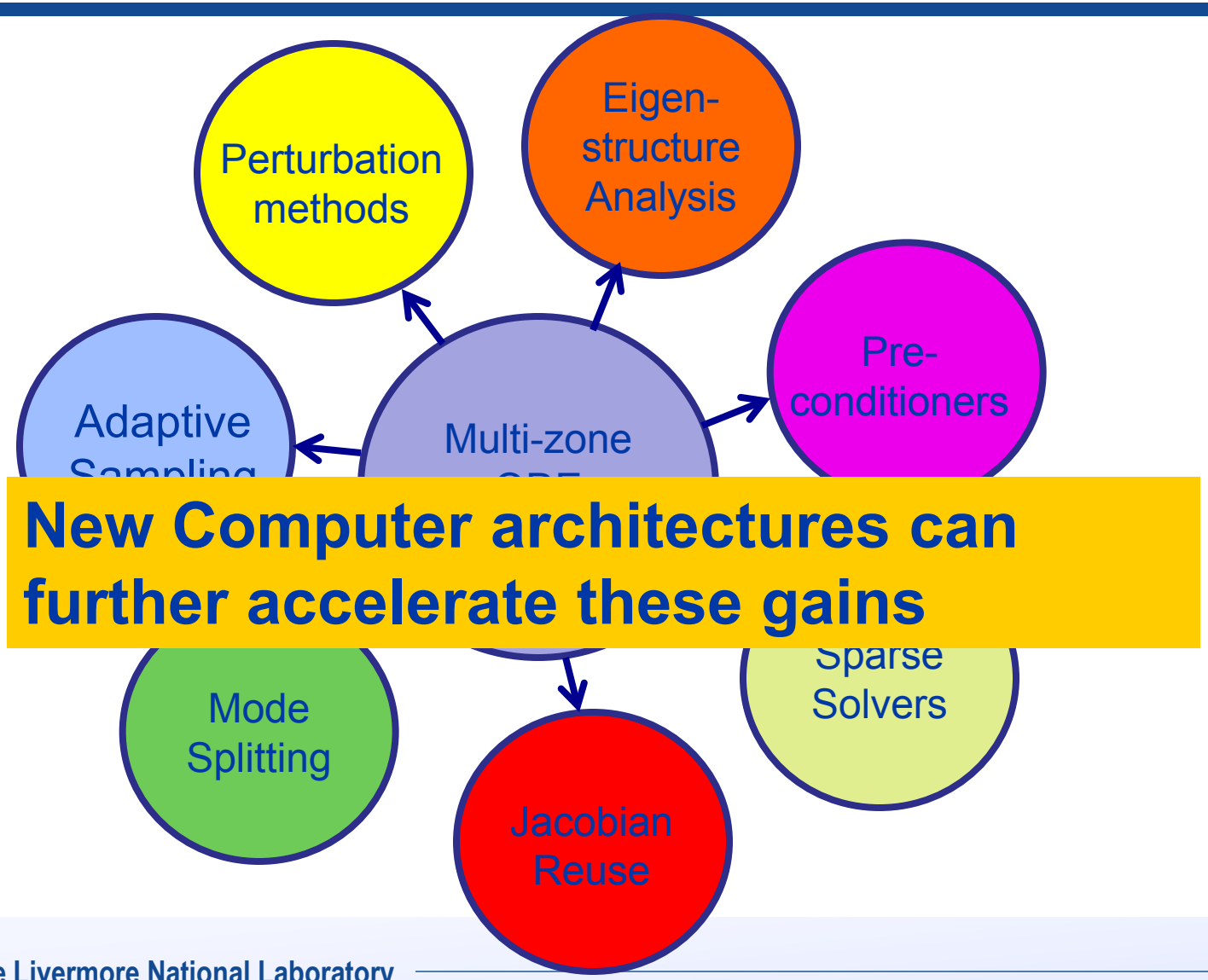
2. New computing architecture

NVIDIA GTX480

3. Advanced numerics



Opportunities for 1000x speedup in computational chemistry cost through applied mathematics



New Computer architectures can further accelerate these gains

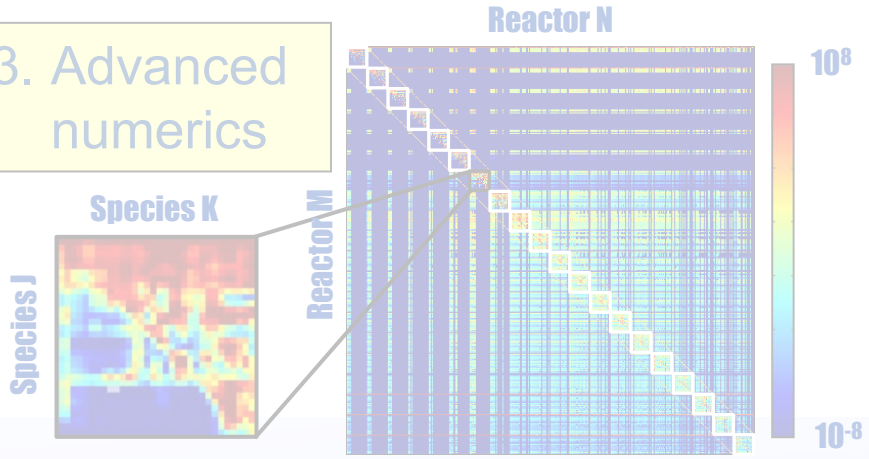
We seek to bring the most physically accurate combustion models for the lowest cost

1. Low-cost models:
   - Mechanism Reduction
   - ANN Ignition Integral

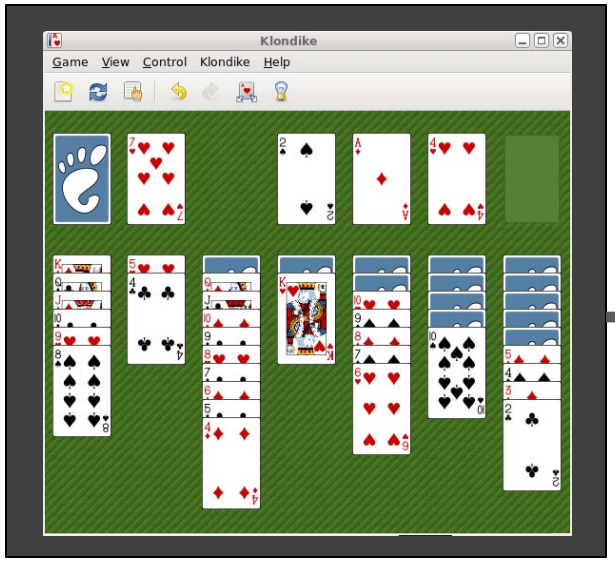
2. New computing architecture

NVIDIA GTX480

3. Advanced numerics



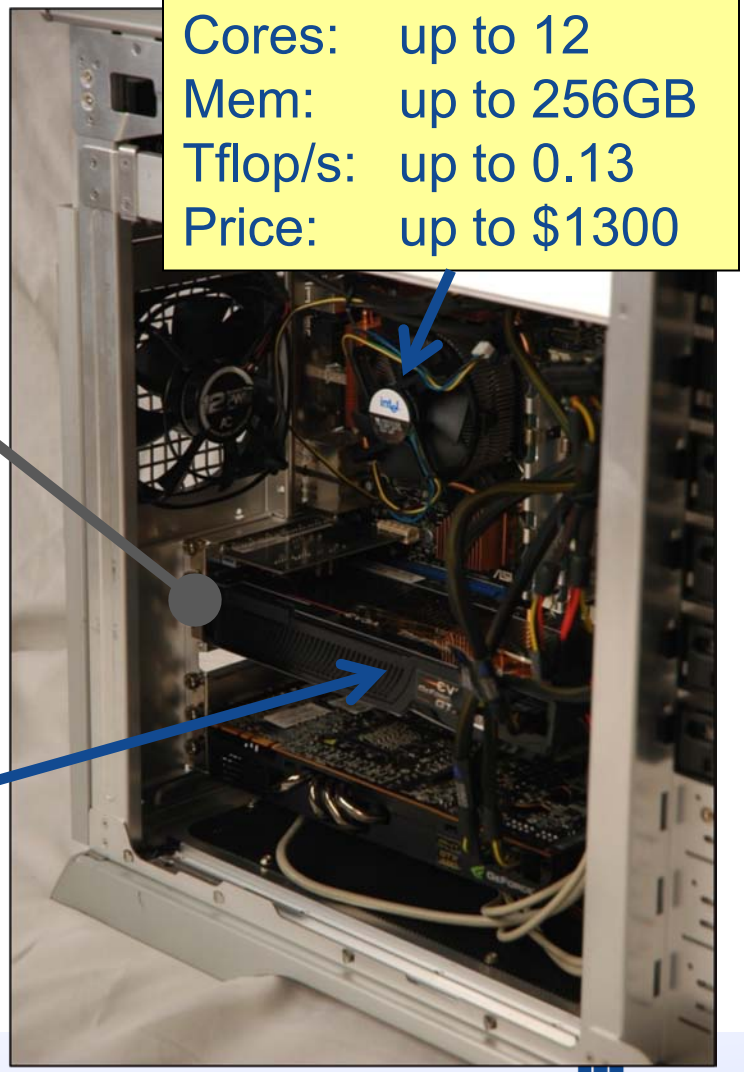
General Purpose Graphical Processing Units (GPGPUs) bring Tflop/s computing power to the desktop



Originally used for graphics intensive applications:

- video games

CPU:	Intel/AMD
Cores:	up to 12
Mem:	up to 256GB
Tflop/s:	up to 0.13
Price:	up to $1300



GPU:	NVIDIA GTX 480
Cores:	480
Mem:	1.5 GB
Tflop/s:	1.35
Price:	$500

NVIDIA's Compute Unified Device Architecture (CUDA) has made computational science on the GPU viable

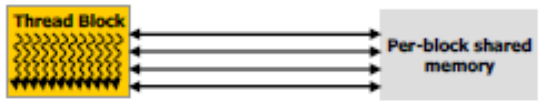
- **GPUs are now programmed with a simple extension to the C language:**
  - New Fermi line offers full C++ support.
  - NVIDIA currently provides free compilers, debuggers and code profilers for all platforms (Linux, Mac and Windows).
  - 3rd party wrappers for most languages (Python, FORTRAN, etc.).
- **Best algorithms have high arithmetic intensity (i.e. many mathematical operations per memory access):**
  - Researchers performing N-body simulations were early adopters (molecular dynamics and astrophysics).
  - Routinely reached +100x speedup.
- **Computational science on the GPUs was in the news recently:**
  - Georgia Tech Research Institute used GPUs to crack passwords.
  - Recommend 12-character random passwords to beat today's GPUs.

GPU architecture and memory controllers require fine-scale parallelism for best performance

Each function (kernel) executes N times for N threads organized in a compute grid of independent thread blocks:

- Images from NVIDIA's CUDA C Programming Guide Version 3.1, 2010.

Optimal GPU algorithms are designed to exploit the fast shared memory



shared (1 - 16)

global (100 - 1600)

texture (<5 - 1600)

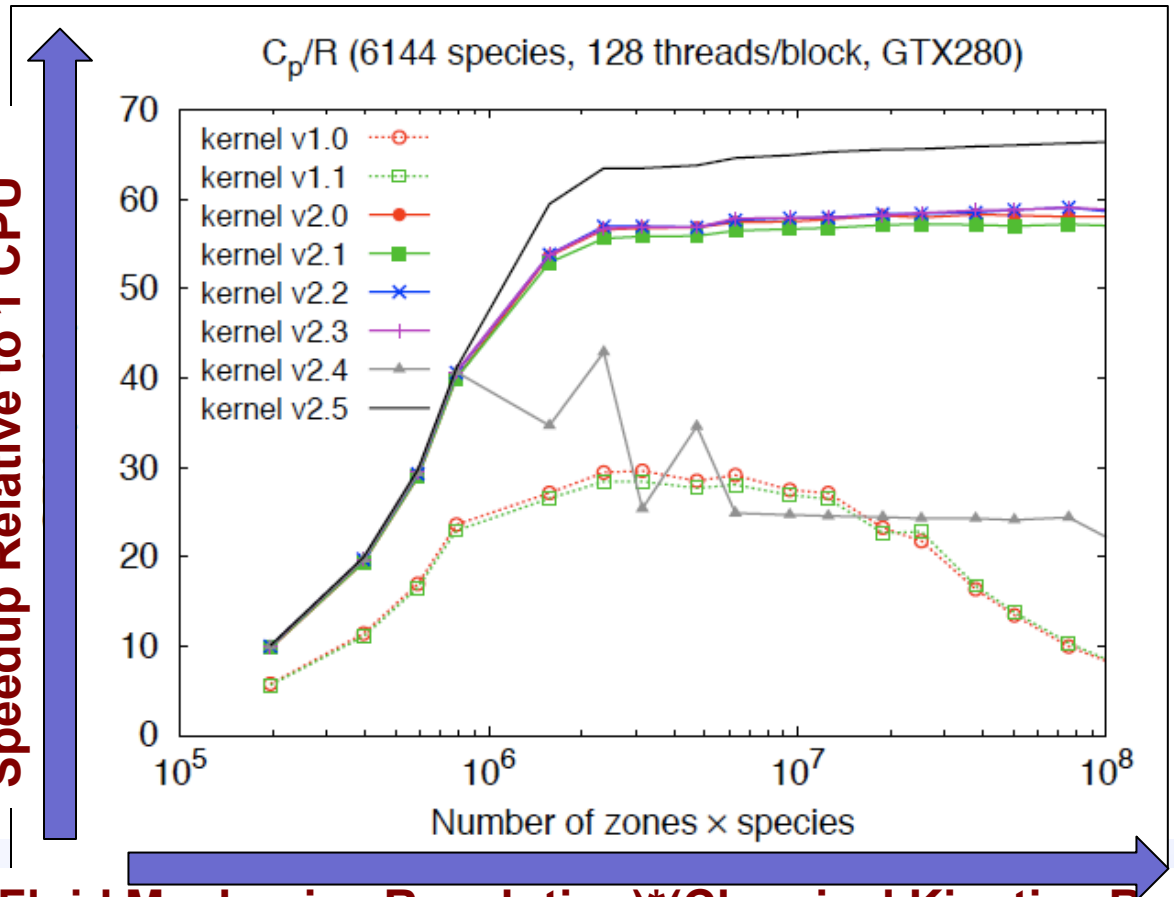
constant (0 - 24) read-only

Evaluation of thermodynamic properties highlights the algorithm design considerations for the GPU

$$\frac{C_p}{R} = \begin{cases} a_1 + a_2\eta + a_3\eta^2 + a_4\eta^3 + a_5\eta^4 & \text{if } \eta \leq \eta_m \\ a_8 + a_9\eta + a_{10}\eta^2 + a_{11}\eta^3 + a_{12}\eta^4 & \text{if } \eta > \eta_m \end{cases}$$

Low Temperature Polynomial (T<1000K)

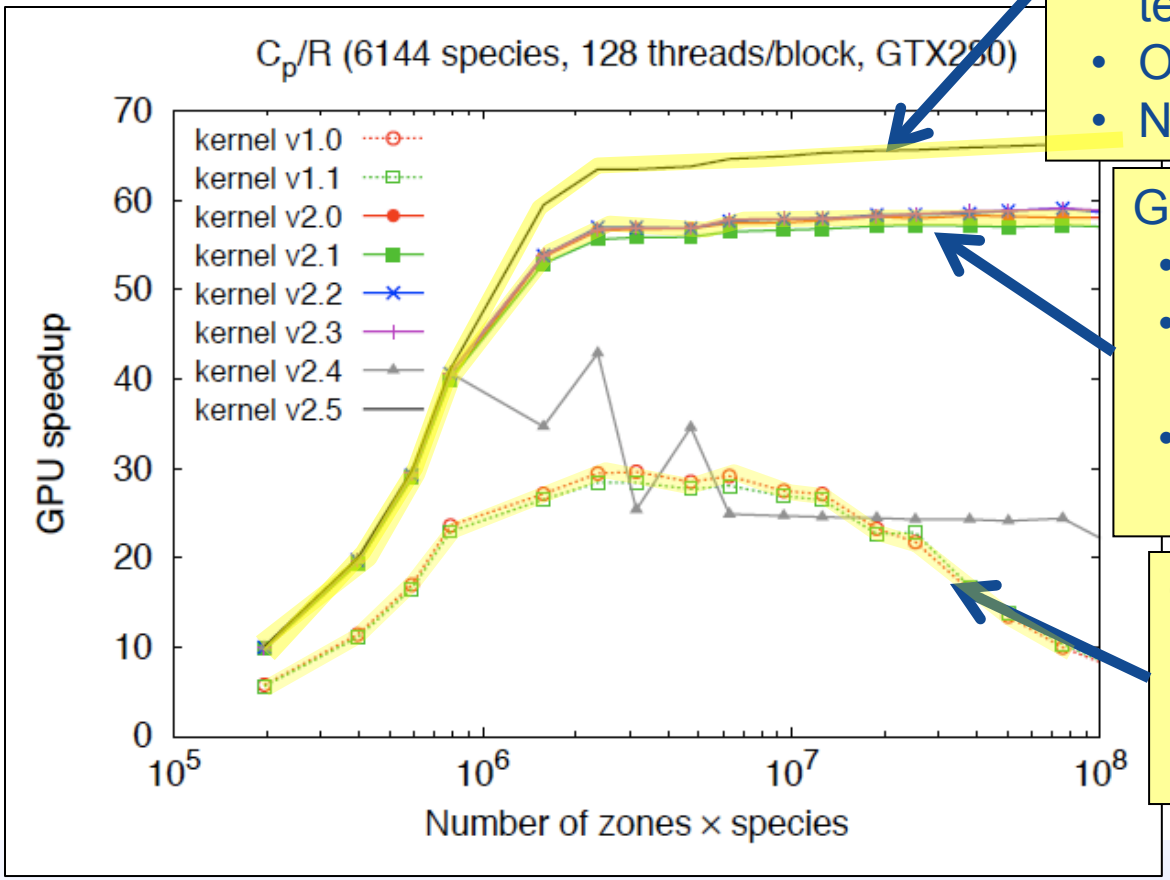
High Temperature Polynomial (T>1000K)



(Fluid Mechanics Resolution)*(Chemical Kinetics Resolution)

Evaluation of thermodynamic properties highlights the algorithm design considerations for the GPU

$$\frac{C_p}{R} = \begin{cases} a_1 + a_2\eta + a_3\eta^2 + a_4\eta^3 + a_5\eta^4 & \text{if } \eta \le \eta_m \\ a_8 + a_9\eta + a_{10}\eta^2 + a_{11}\eta^3 + a_{12}\eta^4 & \text{if } \eta > \eta_m \end{cases}$$

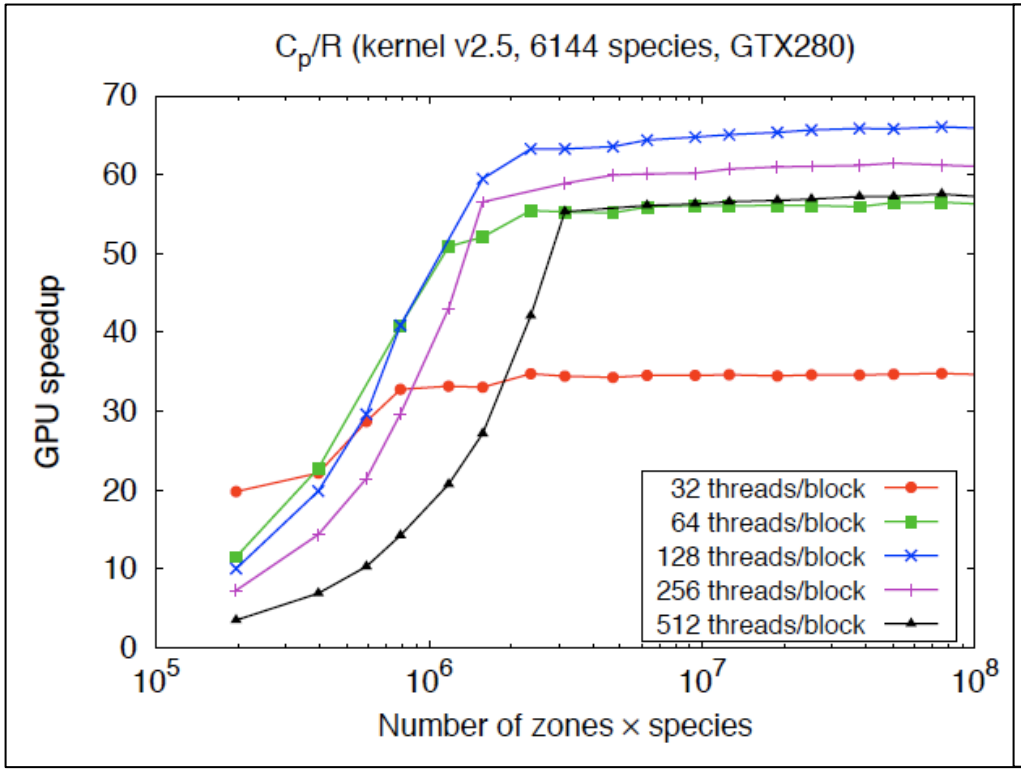


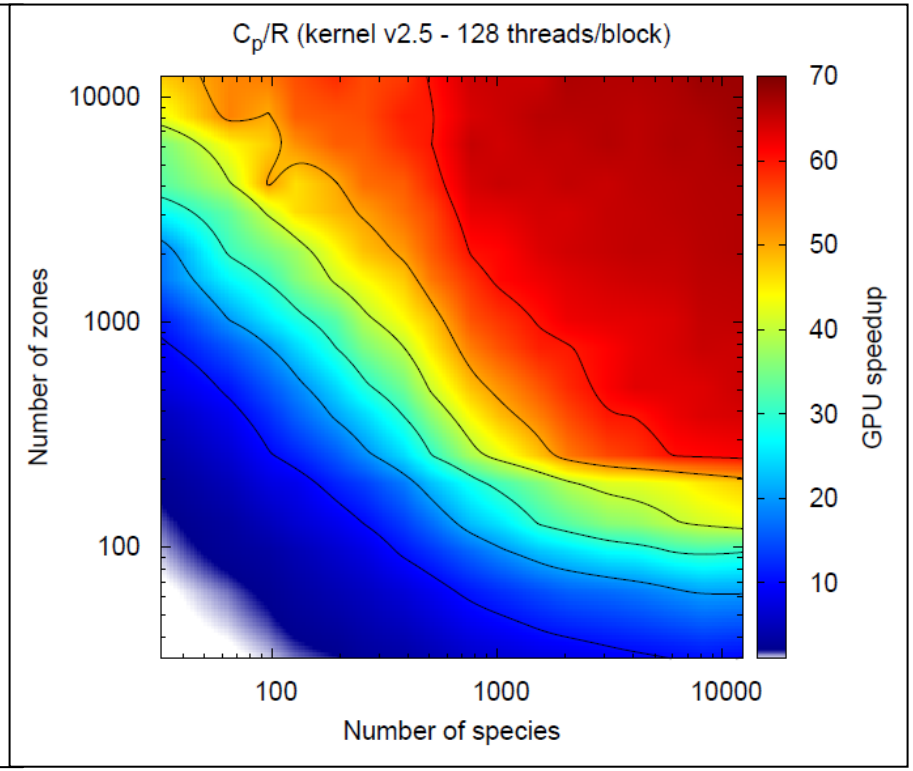
Further optimization
- 65x speedup
- Every thread calculates both temperature branches.
- Only assignment is conditional.
- No thread divergence.

GPU specific coding
- Load a_i in shared memory
- Peak speedup 55x with no size limit on performance
- Load strategies and variable reuse have minor impact

Conventional CPU coding
- No shared memory
- Peak speedup limited to less than 30x

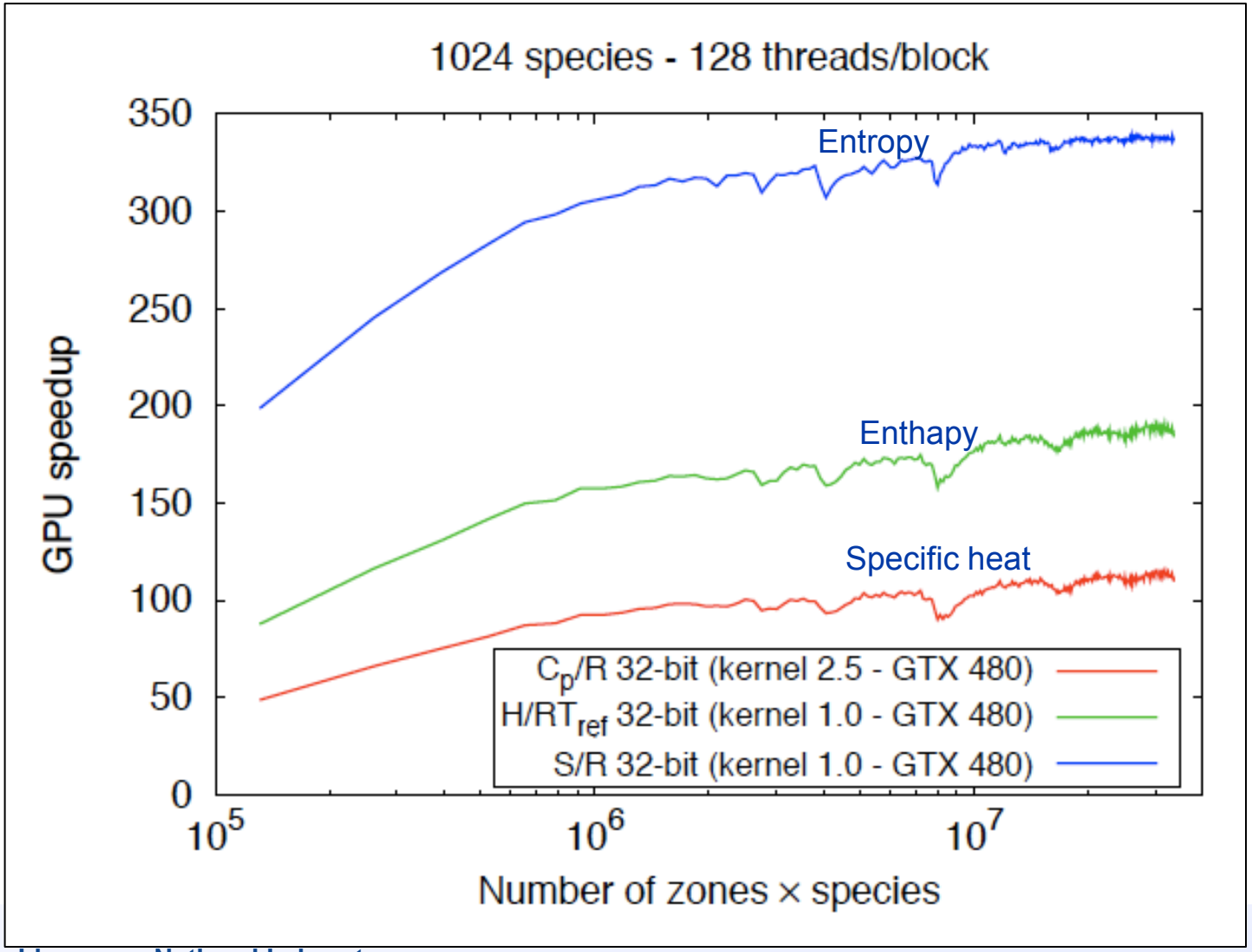
Peak GPU performance is a balance between shared memory reuse and multiprocessor occupancy





- More threads per block: produces greater shared memory reuse between cooperating threads.
- Fewer threads per block: allows more independent blocks to be placed on a multiprocessor effectively hiding memory latency.

Enthalpy and entropy calculations have greater speedup, benefiting from more arithmetic operations per memory access



GPUs are a growth area for bringing scientific computing capability to the desktop

- 64 bit “no-fun” GPU’s being developed for scientific computing
- Democratizing architecture for large-scale computing:
  - Enable greater physics for engineering design
- Puts greater emphasis on programming and numerical methods for effective utilization

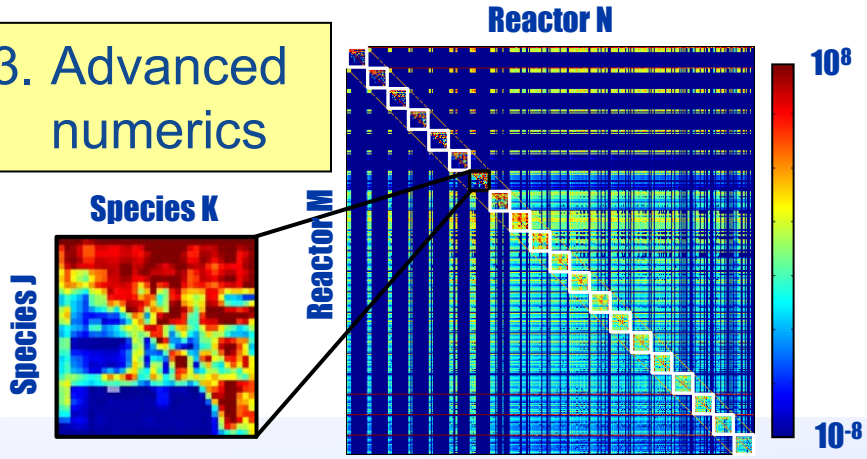
We seek to bring the most physically accurate combustion models for the lowest cost

1. Low-cost models:
   - Mechanism Reduction
   - ANN Ignition Integral

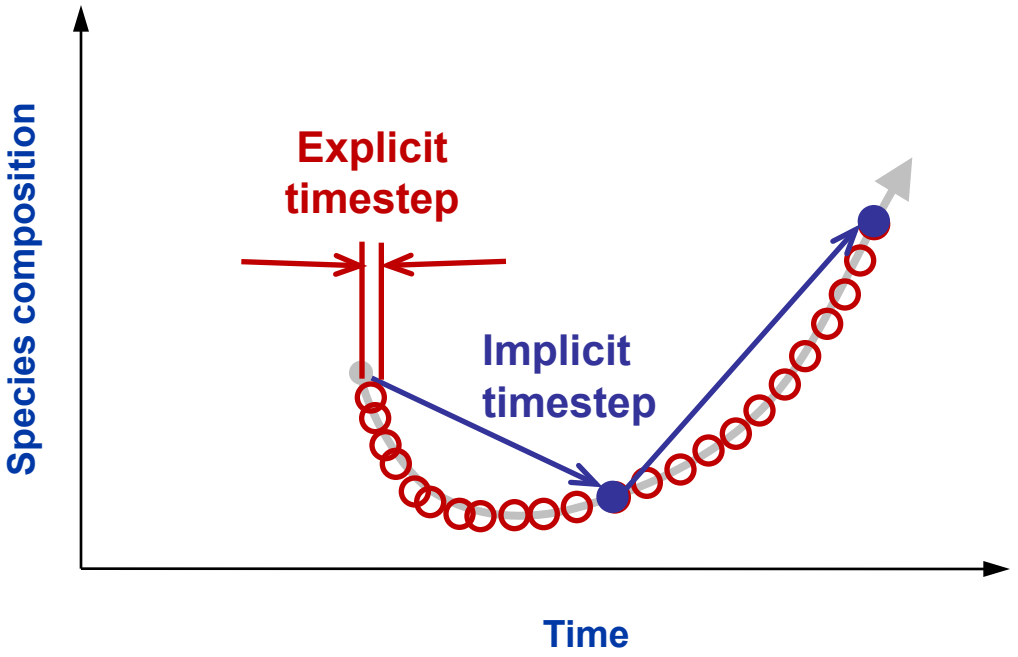
2. New computing architecture

NVIDIA GTX480

3. Advanced numerics



Implicit methods are necessary to integrate the chemical time scales over an engine cycle



Explicit Update
(lower cpu/step)

$$\frac{\partial x_1}{\partial t} = f_1(t, x_1, \ldots, x_N)$$
$$\frac{\partial x_2}{\partial t} = f_2(t, x_1, \ldots, x_N)$$
$$\vdots$$
$$\frac{\partial x_N}{\partial t} = f_N(t, x_1, \ldots, x_N).$$

Implicit Update
(more trajectory data)

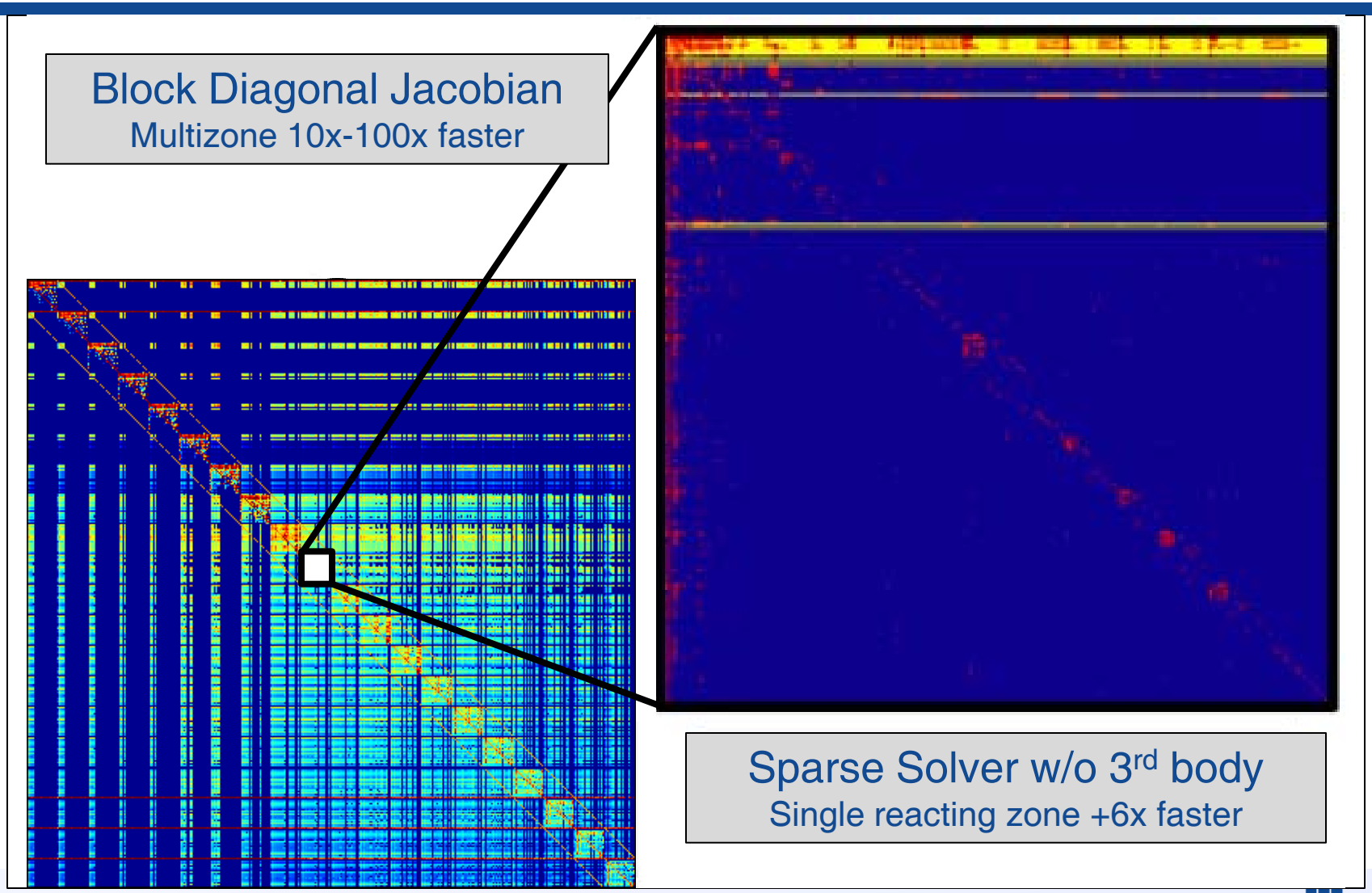
$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & & \frac{\partial f_2}{\partial x_N} \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_N} \end{pmatrix}$$

During ignition:

Δt (explicit) = 10^{-12} to 10^{-15} s

Δt (implicit) = 10^{-6} to 10^{-8} s

Jacobian construction/solution is more than 95% of the simulation cost – a big speedup is possible with smart solvers



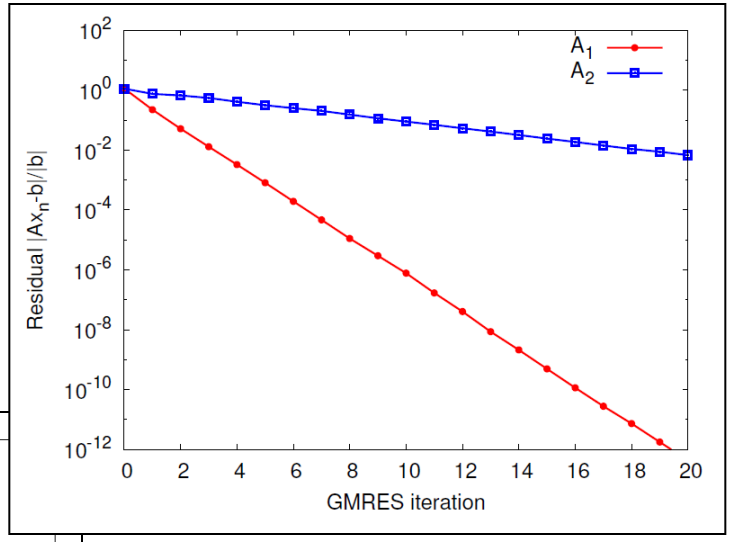
Approximate Jacobians can be used to precondition iterative linear system solvers like GMRES

Generalized Minimal RESiduals

$$E^{(n)} = \frac{\|Ax^{(n)} - b\|_2}{\|b\|_2} \leq \Lambda^n \mathrm{cond}(V)$$

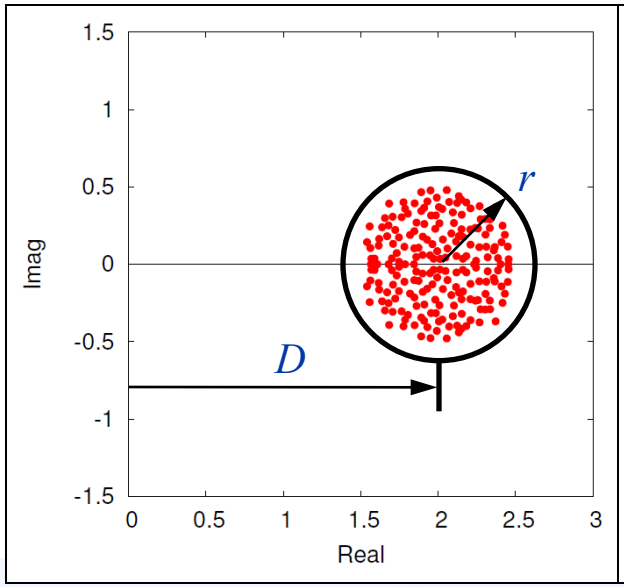
$$\Lambda \approx \frac{r}{D}$$

GMRES Error

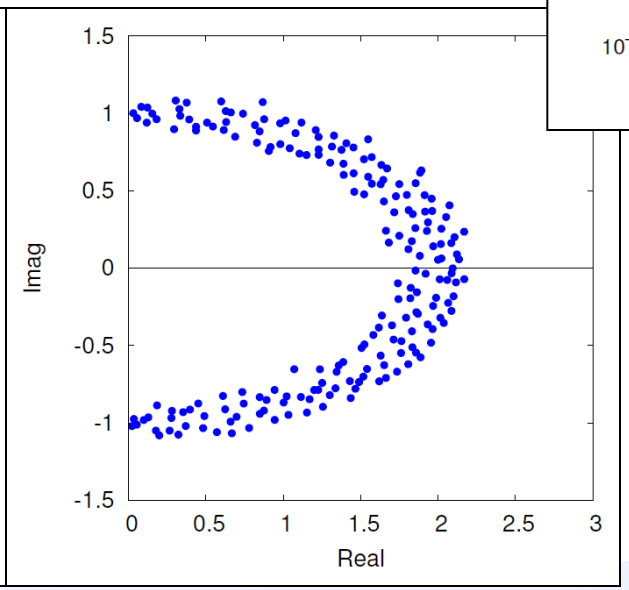


Eigenvalue Spectra (200 x 200)

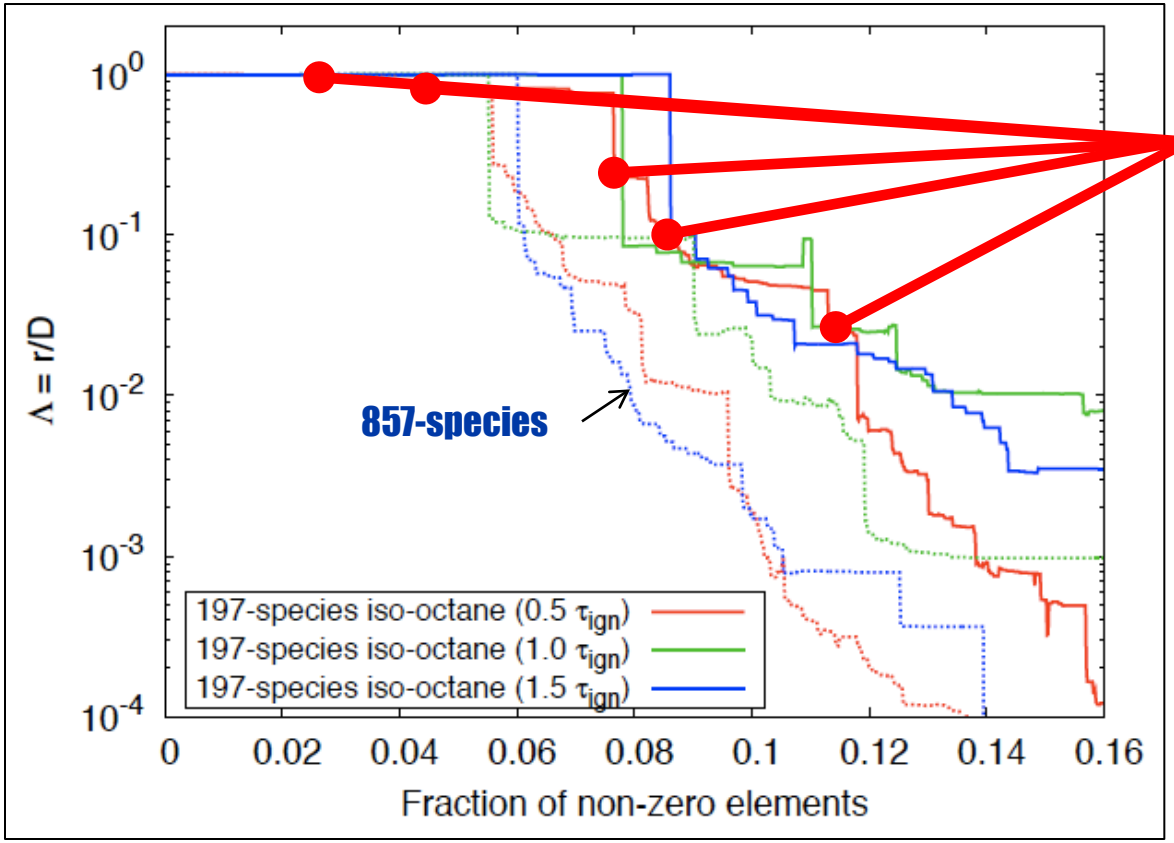
A_1: fast convergence

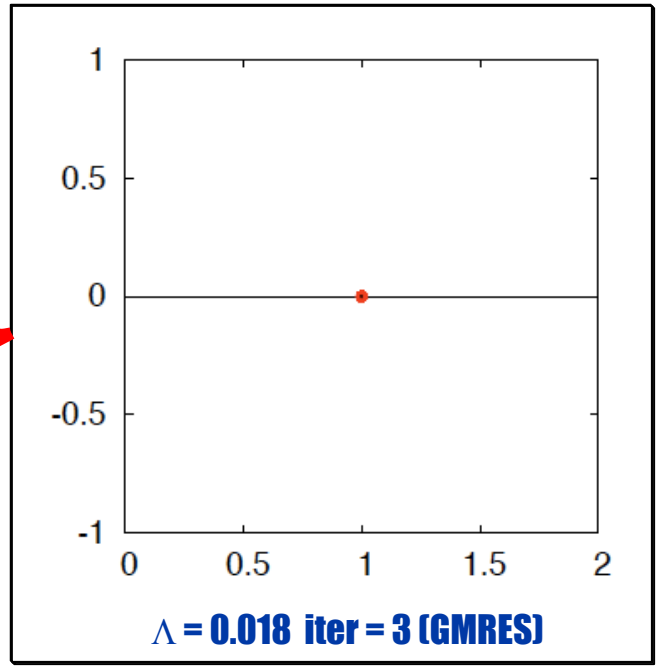


A_2: slow convergence



Direct reaction sorting shows promise to be a general low-cost preconditioner for the Jacobian





Chemical mechanisms continue to evolve, adding more a more compounds that are present in real fuels

- Develop detailed chemical kinetic models for another series iso-alkanes:
  3-methyl alkanes

- Validation of 2-methyl alkanes mechanism with new data from shock tubes, jet-stirred reactors, and counterflow flames
- Develop detailed chemical kinetic models for alkyl aromatics:

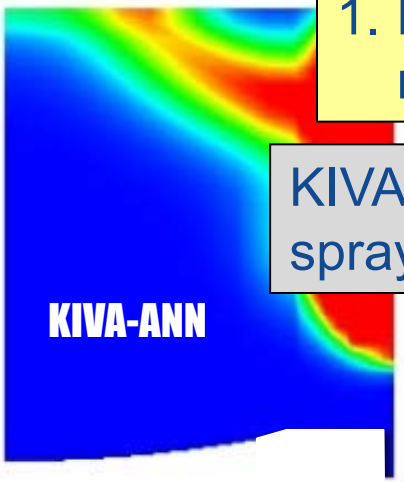
- More accurate surrogates for gasoline and diesel
- Further develop mechanism reduction using functional group method

n-decylbenzene - Diesel Fuels

Continued improvement of physical models and numerical methods will enable utilization of large mechanisms

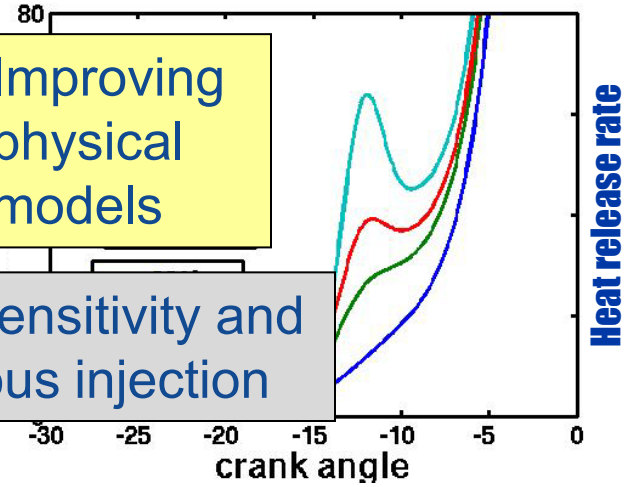
1. Low-cost models

KIVA4+ANN and spray models



2. Improving physical models

LTC sensitivity and gaseous injection



3. New computing architecture

GPU compute tiles for efficient species production rates

NVIDIA GTX480 image:
http://www.nvidia.com/object/product_geforce_gtx_480_us.html

4. Advanced numerics

Combine reaction sort with 2x2 sort