SAND2013-7144C

Uncertainties in *ab initio* rate coefficient calculations

Judit Zádor
Sandia National Laboratories

The VIIIth Congress of the International Society of Theoretical Chemical Physics
April 25 – 31, 2013
Budapest, Hungary

Contributors and support

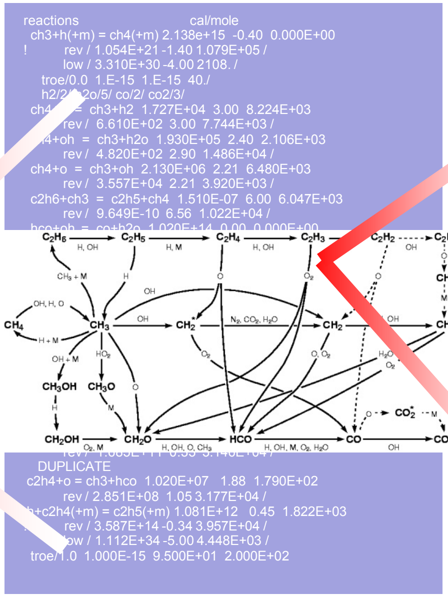
Habib N. Najm
Jens Prager

Ewa Papajak

Division of Chemical Sciences, Geosciences, and Biosciences, the Office of Basic Energy Sciences, the U.S. Department of Energy under DOE Contract Numbers DE-AC04-94AL85000

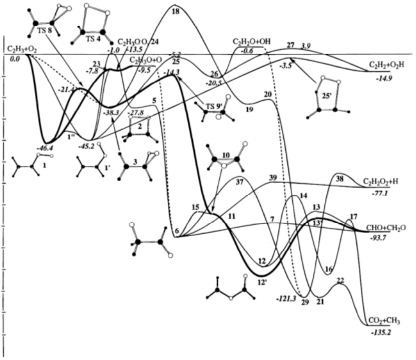
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the National Nuclear Security Administration.

The hierarchy of combustion chemistry



```
reactions                          cal/mole
  ch3+h(+m) = ch4(+m) 2.138e+15  -0.40  0.000E+00
!         rev / 1.054E+21 -1.40 1.079E+05 /
          low / 3.310E+30 -4.00 2108. /
    troe/0.0  1.E-15  1.E-15  40./
    h2/2/  2o/5/ co/2/ co2/3/
  ch4     =  ch3+h2  1.727E+04  3.00  8.224E+03
          rev /  6.610E+02  3.00  7.744E+03 /
   4+oh  =  ch3+h2o  1.930E+05  2.40  2.106E+03
          rev /  4.820E+02  2.90  1.486E+04 /
  ch4+o  =  ch3+oh  2.130E+06  2.21  6.480E+03
          rev /  3.557E+04  2.21  3.920E+03 /
  c2h6+ch3  =  c2h5+ch4  1.510E-07  6.00  6.047E+03
          rev /  9.649E-10  6.56  1.022E+04 /
  hco+oh  =  co+h2o  1.020E+14  0.00  0.000E+00

    DUPLICATE
  c2h4+o = ch3+hco  1.020E+07   1.88  1.790E+02
          rev / 2.851E+08  1.05 3.177E+04 /
 h+c2h4(+m) = c2h5(+m) 1.081E+12   0.45  1.822E+03
          rev / 3.587E+14 -0.34 3.957E+04 /
          ow / 1.112E+34 -5.00 4.448E+03 /
   troe/1.0  1.000E-15  9.500E+01  2.000E+02
```



Combustion involves the interaction of fluid flow and chemistry

Complex chemical mechanisms

- thousands of pressure- and temperature-dependent rate coefficients
- increasingly rely on theory
- mechanisms are validated in comparison to bulk experiments
- uncertainty tools are widely used

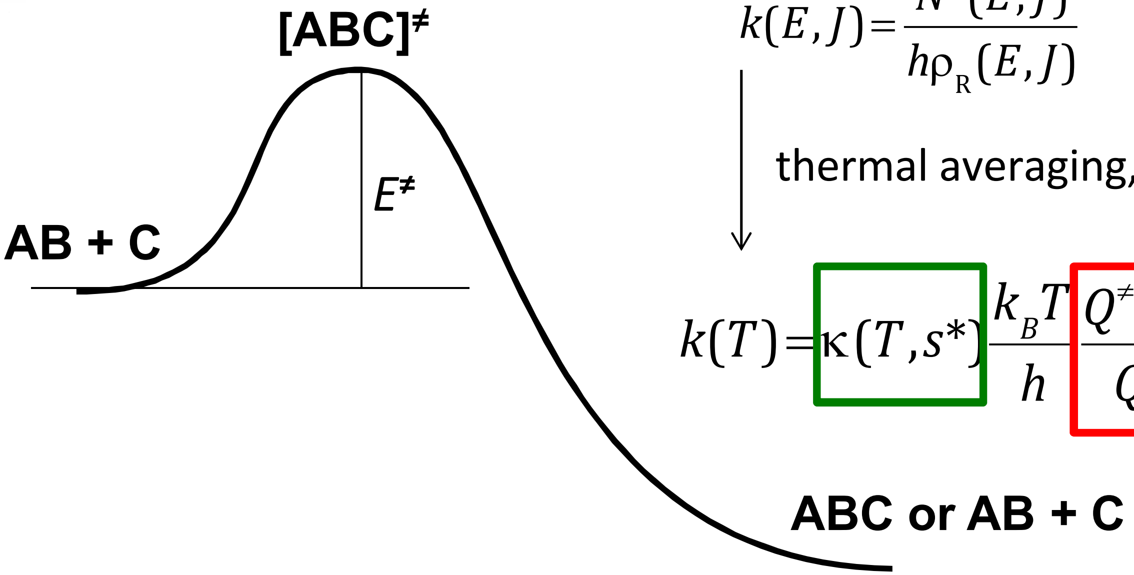
Theoretical chemical kinetics

- rate coefficients are calculated from statistical theory based on microscopic parameters
- rate coefficients are tested against elementary rate coefficient measurements
- uncertainty analysis tools are rarely used

Similar can be said about atmospheric chemistry.

$$\frac{\text{KINETICS}}{\text{MECHANISM}} = \frac{\text{FACT}}{\text{FICTION}}$$

Sources of uncertainty in the high-pressure limit of association reactions with barriers or simple abstraction reactions



$$k(E,J)=\frac{N^{\neq}(E,J)}{h\rho_{\mathrm{R}}(E,J)}$$

$\downarrow$ thermal averaging, canonical variational TST

$$k(T)=\boxed{\kappa(T,s^*)}\,\frac{k_BT}{h}\,\boxed{\frac{Q^{\neq}(T,s^*)}{Q_R(T)}}\,\boxed{\exp(-E^{\neq}(s^*)/RT)}$$

*s**: variationally determined, 3*N*-7 dimensional dividing surface, for reactions with a barrier often well approximated by a ridge including the saddle point(s) in question

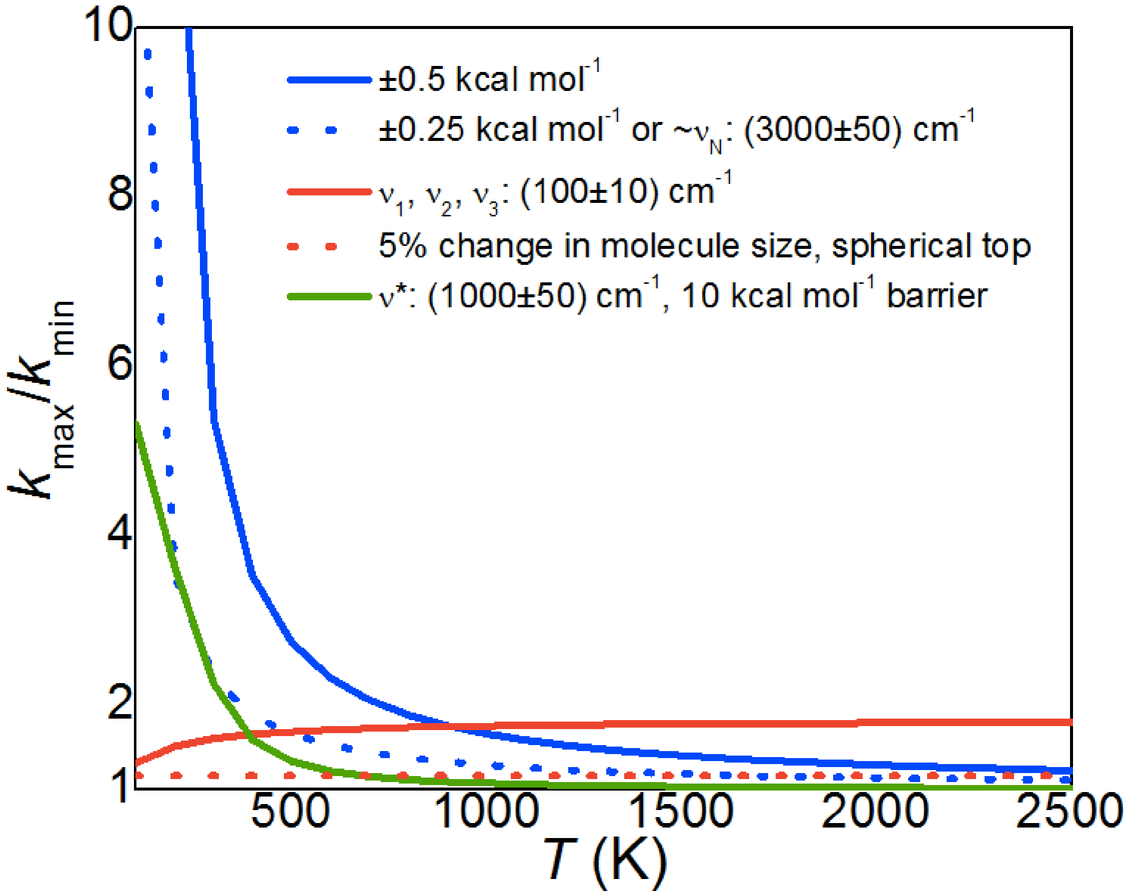
***E*≠**: barrier height, E_{el} + *ZPE*

Q: electronic, rotational and vibrational (usually largely uncoupled) partition function

κ: tunneling correction

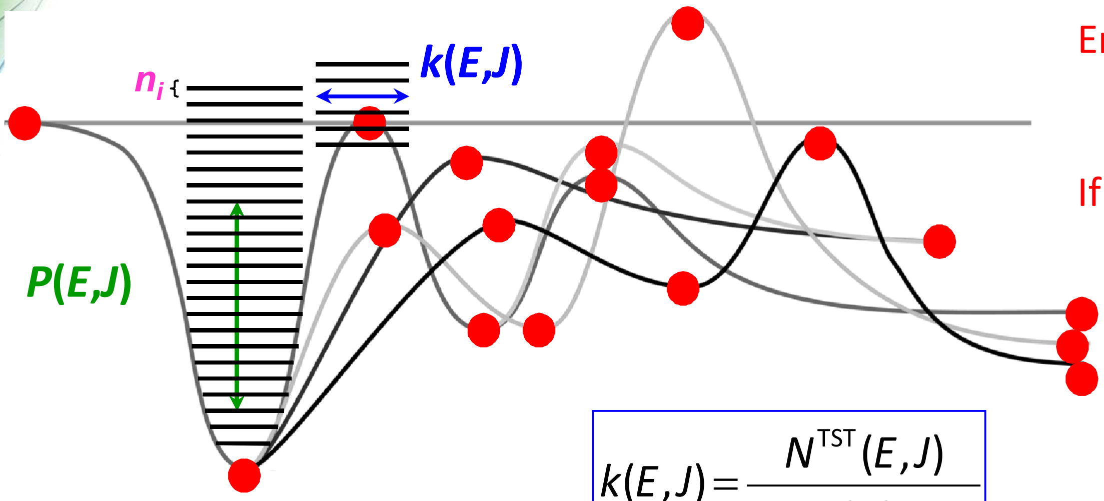
Sources of uncertainty in the high-pressure limit of association reactions with barriers or simple abstraction reactions

$$k(T)=\boxed{\kappa(T,s^*)}\frac{k_BT}{h}\boxed{\frac{Q^{\neq}(T,s^*)}{Q_R(T)}}\boxed{\exp(-E^{\neq}(s^*)/RT)}$$



What are the actual errors, and how much can we rely on cancellation of errors?

Thermal rate coefficients are highly averaged macroscopic quantities, both fundamentally and pragmatically



$$k(E,J) = \frac{N^{TST}(E,J)}{h\rho^{reactant}(E,J)}$$

E.g. *n*-butyl + O_2 non-variational calculation:
17 x 5 x (3 x 15 – 2) + 17 x 3 = 3706 fundamental parameters**.**

➔ A master equation calculation will extract S_{stable} x (S_{stable} - 1) rate coefficients = 72 rate coefficients.

Experimentally typically a few can be determined.

Energy and other properties of stationary points are calculated from quantum chemistry.

If *S* structures in *N*-atom system, each has *C* conformers:
C x *S* energies,
C x *S* x (3*N* – 6) harmonic frequencies,
3*C* x *S* rotational constants
= *C* x *S* x (3*N* – 2) parameters

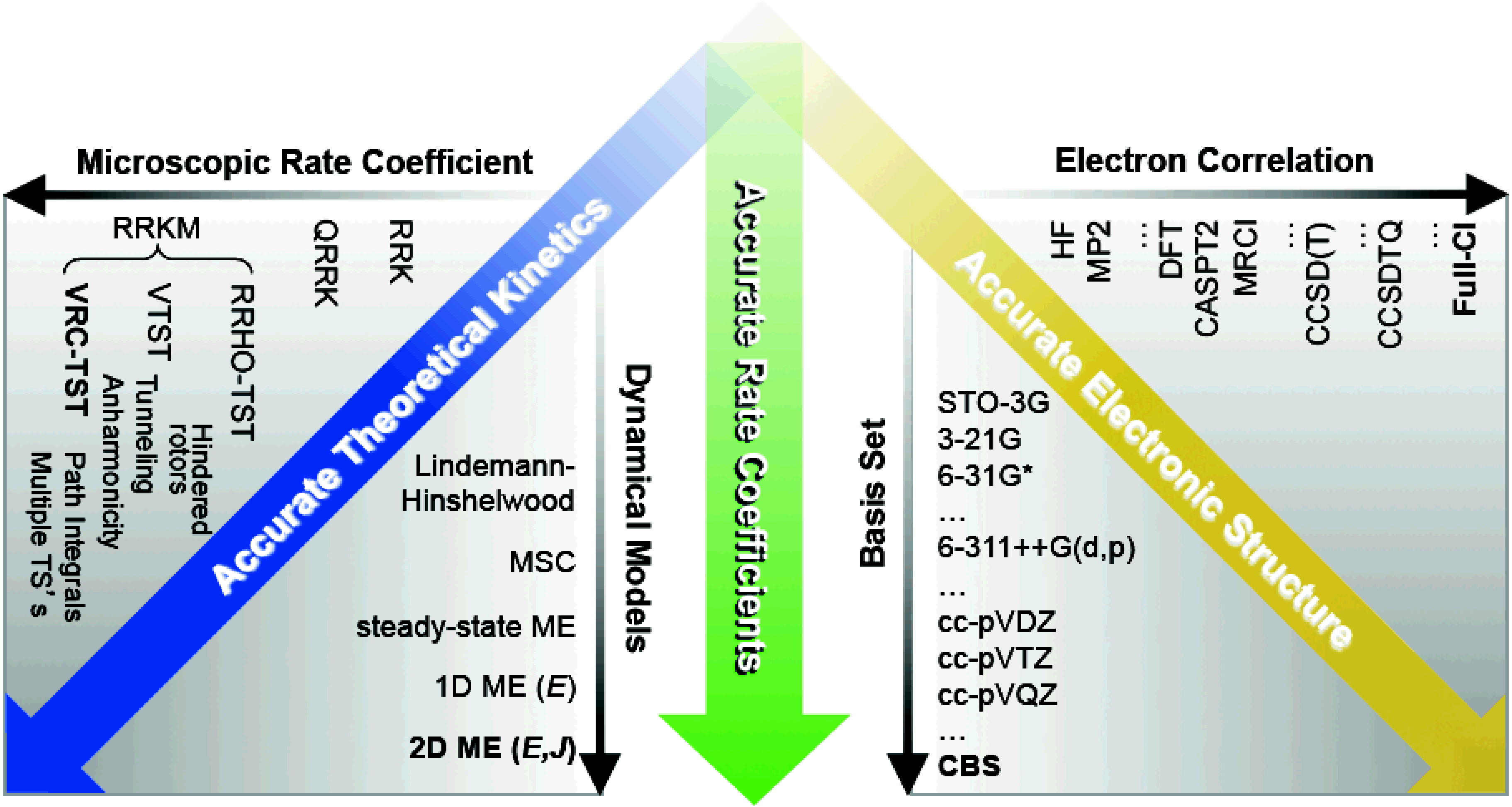
Transition-state theory (TST) is used to compute energy and angular-momentum specific rate constants, *k*(*E*,*J*).

If variational TST is required, number of parameters is multiplied by ~10-1000

Collisional energy transfer redistributes population among energy “levels”.

~ 3 x S parameters, or more…

The path to accurate theoretical kinetics calculations



Microscopic Rate Coefficient

RRKM
- RRHO-TST
- VTST
- VRC-TST

QRRK

RRK

Hindered rotors

Tunneling

Anharmonicity

Path Integrals

Multiple TS's

Dynamical Models
- Lindemann-Hinshelwood
- MSC
- steady-state ME
- 1D ME (*E*)
- **2D ME (*E*,*J*)**

Accurate Theoretical Kinetics

Accurate Rate Coefficients

Accurate Electronic Structure

Electron Correlation
- HF
- MP2
- …
- DFT
- CASPT2
- MRCI
- …
- CCSD(T)
- …
- CCSDTQ
- …
- **Full-CI**

Basis Set
- STO-3G
- 3-21G
- 6-31G*
- …
- 6-311++G(d,p)
- …
- cc-pVDZ
- cc-pVTZ
- cc-pVQZ
- …
- **CBS**

The errors in various parts of the calculations need to be balanced.

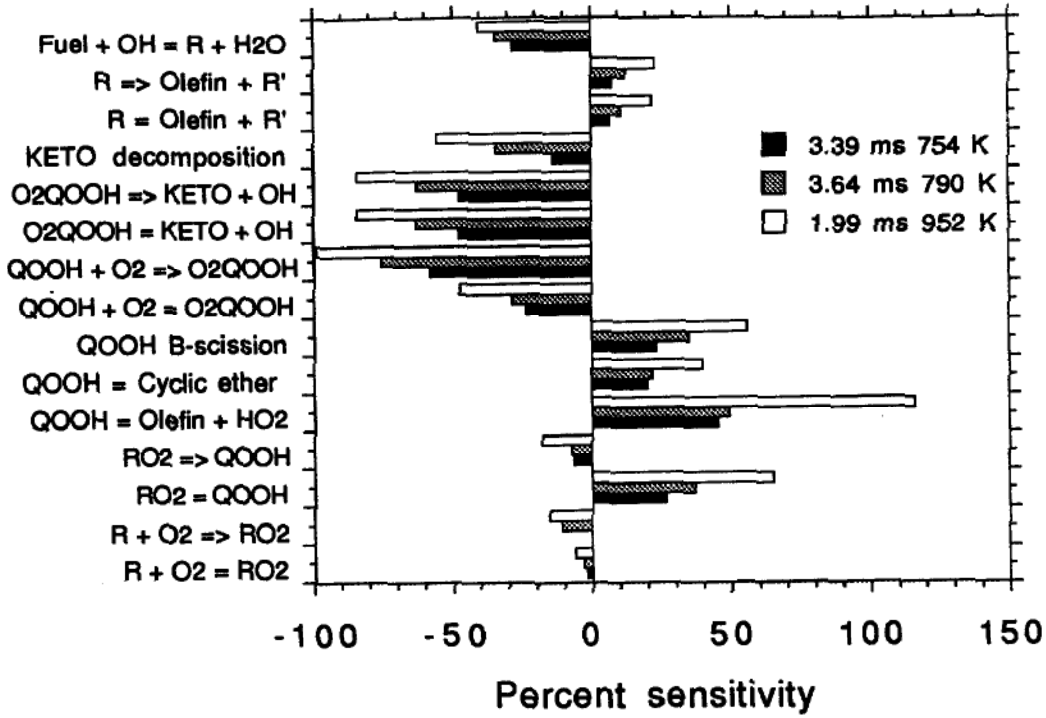
What is the required accuracy?

Chemical accuracy: <1 kcal mol^{-1} uncertainty (or rather 0.5 kcal mol^{-1})

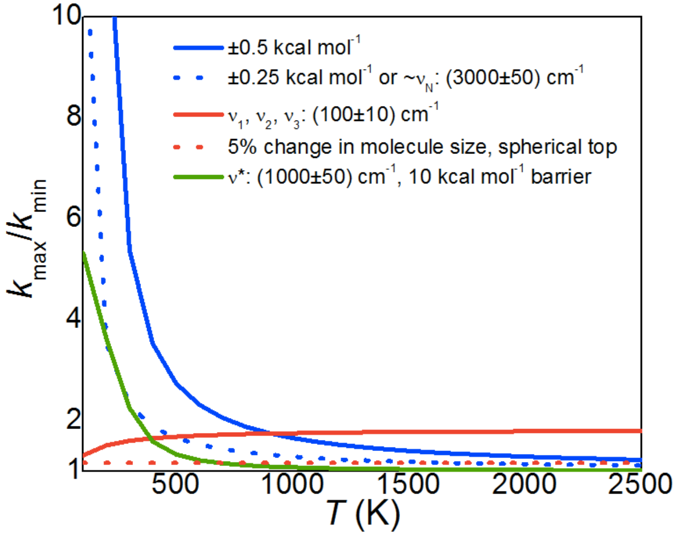
Spectroscopic accuracy: 1 cm^{-1}, or 15 cm^{-1} and 0.005 A

Peterson et al., 2012, Theor Chem Acc

Kinetic accuracy: < 2x uncertainty in the rate coefficient

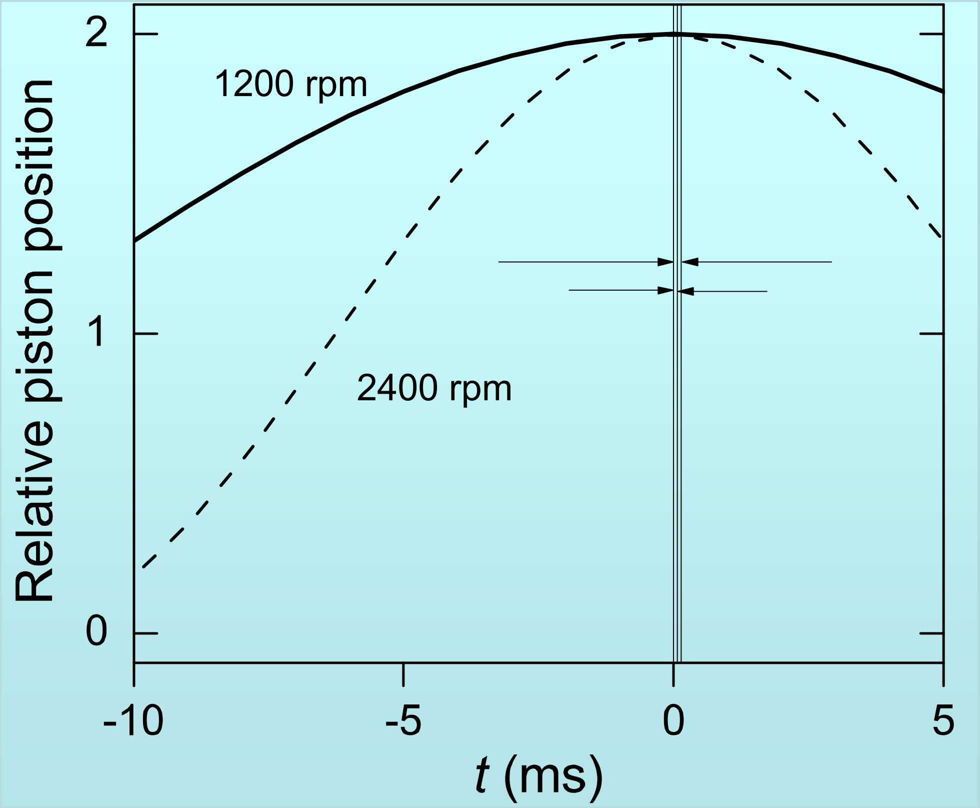


sensitivities of ignition delay time to a 2x increase in rate coefficients



Curran et al. LLNL *n*-heptane mechanism

HCCI engine operation requires the precise knowledge of autoignition delay times

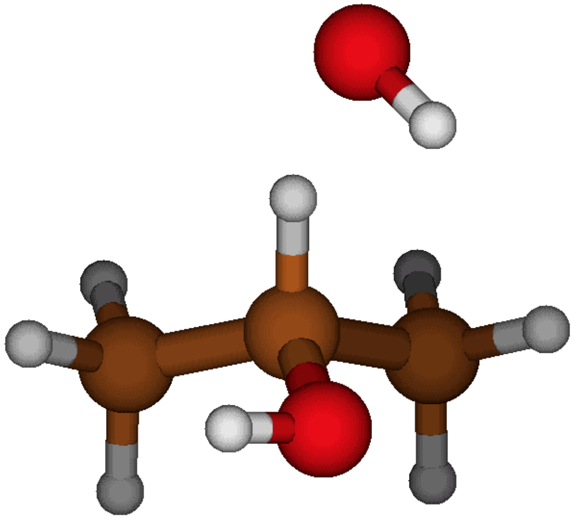


Ignition has to be timed at a minimum of 1 crank angle precision

@ 1200 rpm ~ 140 µs
@ 2400 rpm ~ 70 µs

Typical ignition time in HCCI: 2-4 ms ➔ Ignition delay predictions within ~3%.

The Bayesian case study on the isopropanol + OH rate coefficient calculation



The theoretical framework
Electronic structure calculations

Various levels of electronic structure calculations were used to calculate geometries, energies and frequencies

- QCISD(T)/CBS//B3LYP/6-311++G(d,p)
- QCISD(T)/CBS//M06-2X/6-311++G(d,p)
- CCSD(T)-F12/cc-PVTZ-F12//M06-2X/6-311++G(d,p)
- QCISD(T)/CBS//MP2/6-311++G(d,p)
- QCISD(T)/CBS//CASPT2(3e,3o)/aug-cc-pV*n*Z (*n*=D and T)
- CBS-QB3 MUD = 0.69 kcal mol^{-1}
- G3B3
- G3MP2

$\left.\begin{array}{l}\text{G3B3}\\ \text{G3MP2}\end{array}\right\}$ MUD ~ 1.1 kcal mol^{-1}

T1~ 0.012

CBS extrapolation is done with the formula of Martin and Feller

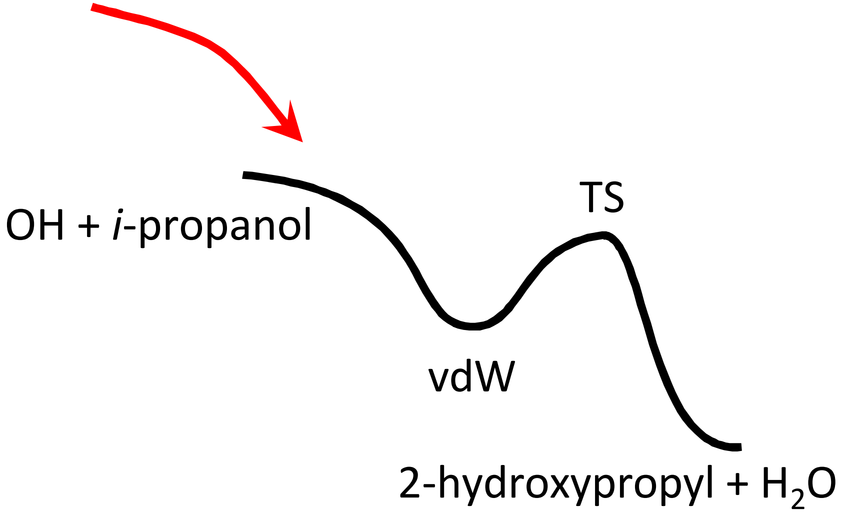
$$E_{\infty} = E_{l_{\max}} - B/(l_{\max}+1)^4$$

The uncertainty of these methods is known in principle, but

- it is just a statistical measure (MAD, MUD)
    - → there are always outliers in test sets
- best for stable species, to a lesser extent for radicals, and definitely not for transition states
- for non-composite methods geometry optimization choice is important
    - → low-frequency modes
    - → ZPE

The theoretical framework
Microscopic rate coefficient calculations

- Microcanonical Rice-Ramsperger-Kassel-Markus (RRKM) theory was used for calculating *k*(*E*)
- RRHO + hindering potentials, corrected for the coupled OH modes at TS
- Asymmetric Eckart barrier for tunneling probabilities
- Phase-space theory for barrierless entrance
- Spin-orbit coupling of the OH radical



Overview of most uncertain calculated molecular properties

			(kcal/mol)		(cm^{-1})			
No.		Method	E+ ΔZPE$_H$	ΔZPE$_H$	ν_2	ν_{27}	ν^*	$\Delta d\%$
1	QCISD(T)/CBS	B3LYP	−1.36	+1.15[a]	143	2663	129i	1.68
2		M062x	−1.18	−0.48	111	1912	588i	4.52
3		MP2	−1.94	−0.93	88	1571	832i	5.32
4		CASPT2/ADZ	−1.18	−0.42	99	1774	624i	4.33
5		CASPT2/ATZ	−1.16	−0.42	94	1797	598i	4.39
6		CBS-QB3	−1.19	+1.08	161	2534	135i	2.33
7		G3B3	−0.41	+0.94	134	2560	116i	2.37
8		G3MP2	−1.89	−1.89	106	1647	2929i	7.79
9		MP2/6-31G* [30]	−1.14	−1.26	–	–	1583i	–
10		BHandHLYP [31]	−0.86[b]	–	–	–	732.4i	–

[a] The anharmonic ΔZPE is +1.33.

[b] CCSD(T)/6-311G(d,p).

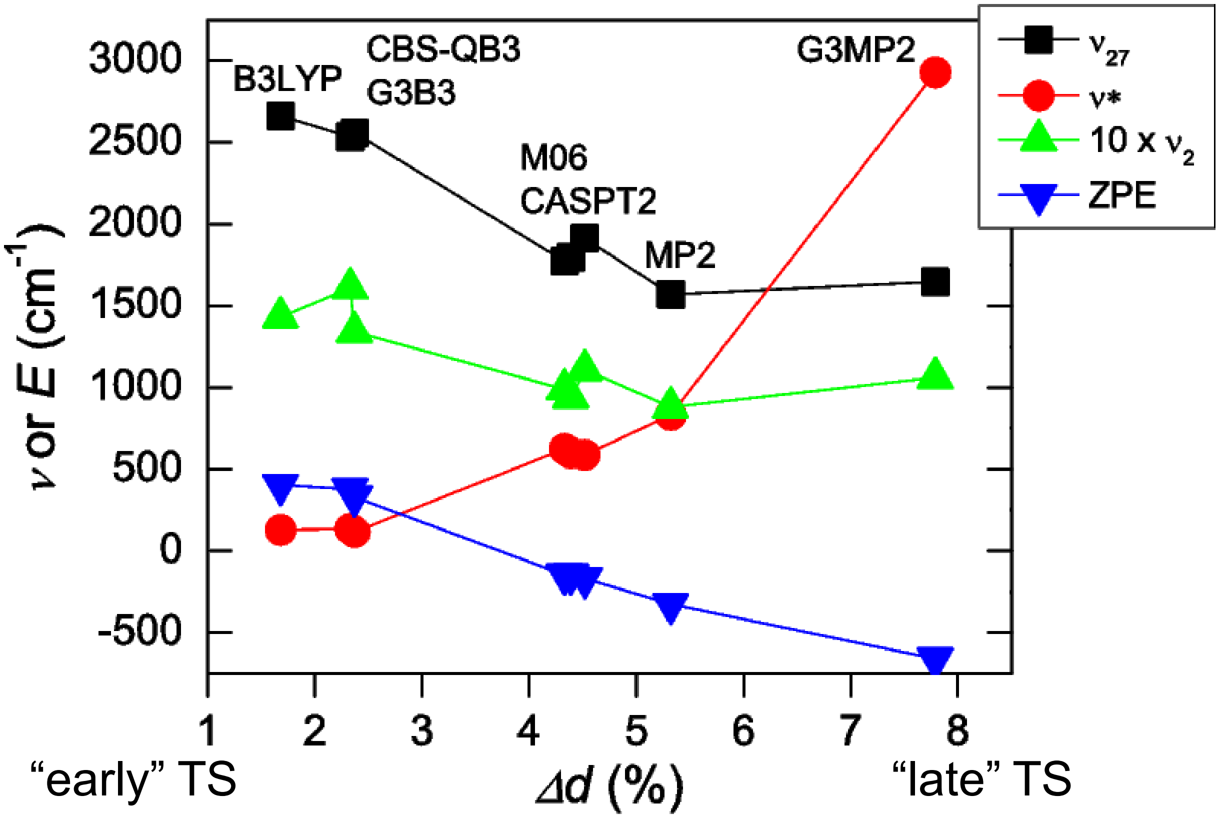
CCSD(T)-F12/cc-pVTZ-F12//M06-2X/6-311++G(d,p): -1.83 kcal mol^{-1}

Most important uncertain parameters: E, ν_2, ν^*

Considerable scatter in barrier height E (E_{el}+ΔZPE): -0.41 to -1.94 kcal mol^{-1}

Largest variability in imaginary frequency ν^* at saddle point

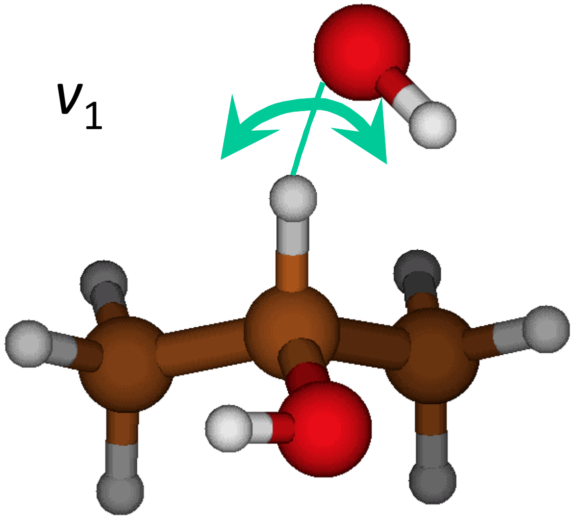
Calculated molecular properties are correlated



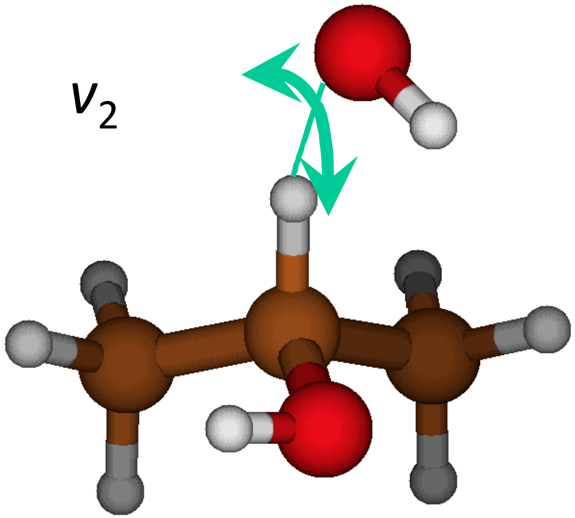
- Different effects on rate coefficients:
  (a) compensating or
  (b) amplifying
- Correlations are important for forward UQ
- Variation in properties of the reactants are negligible

Uncertain normal mode frequencies at the transition state

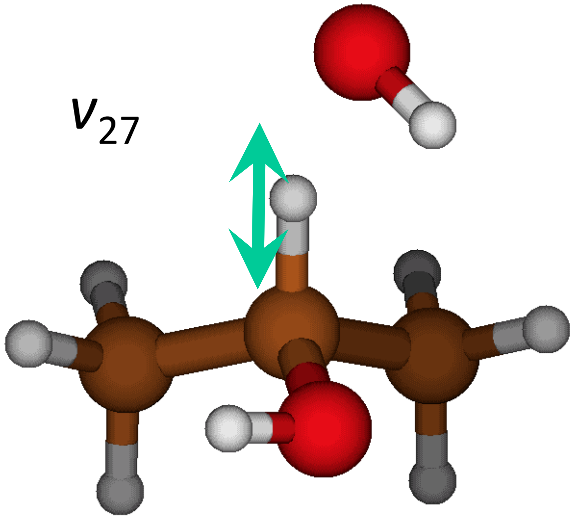
ν_1



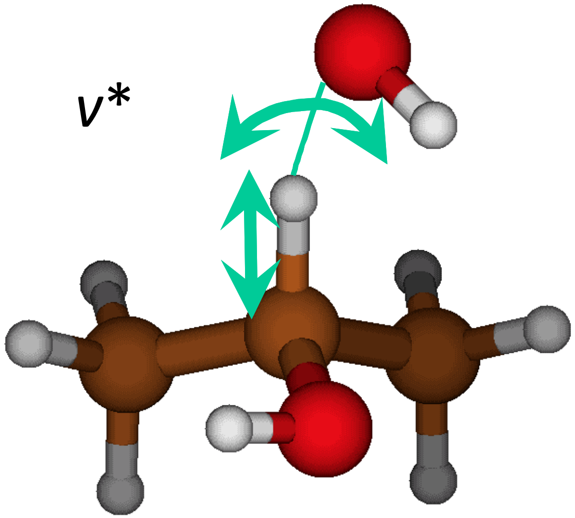
ν_2



ν_{27}

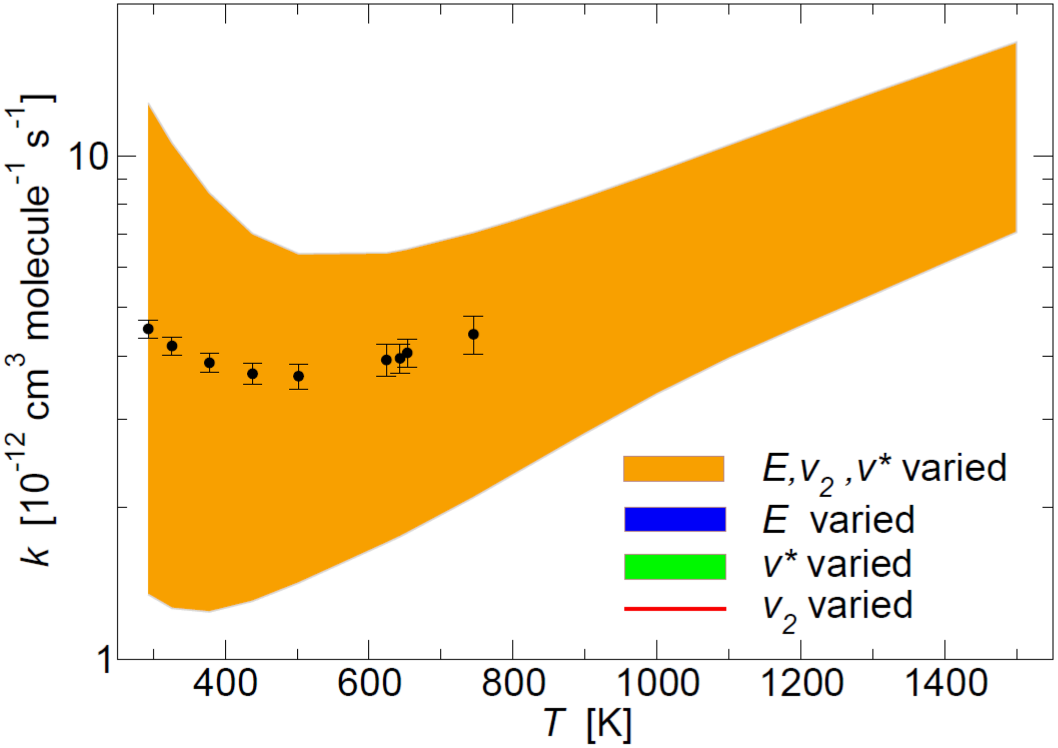


ν^*



Uncertainty propagation using uncorrelated QC parameters:

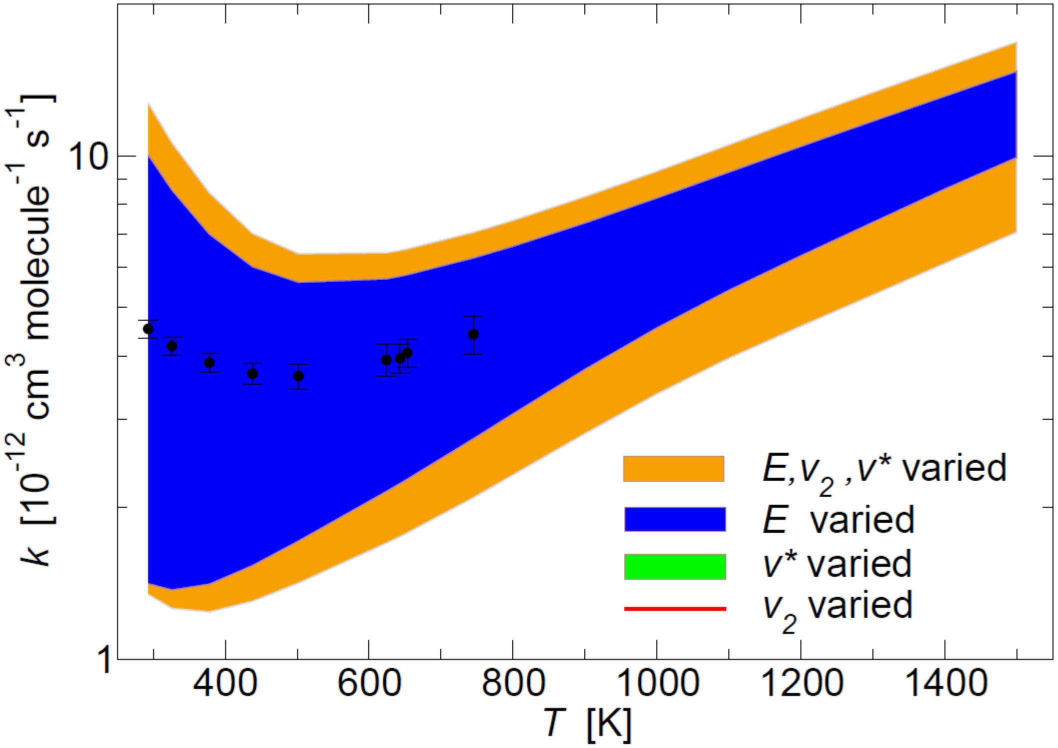
Ranges: E : -2.0...-0.7 kcal mol^{-1}, v_2 : 70...150 cm^{-1}, v^* : 50...850 cm^{-1}



- 95% intervals
- Independent, uniform distribution
- Ratio of upper to lower bound:
  - (a) 10 at 300K
  - (b) 2.5 at 800K
- *E* dominates up to 800K
- v_2 small temperature dependence
- v^* minor importance, only to 500K
- (R1) is challenging from theoretical/computational aspect

Uncertainty propagation using uncorrelated QC parameters:

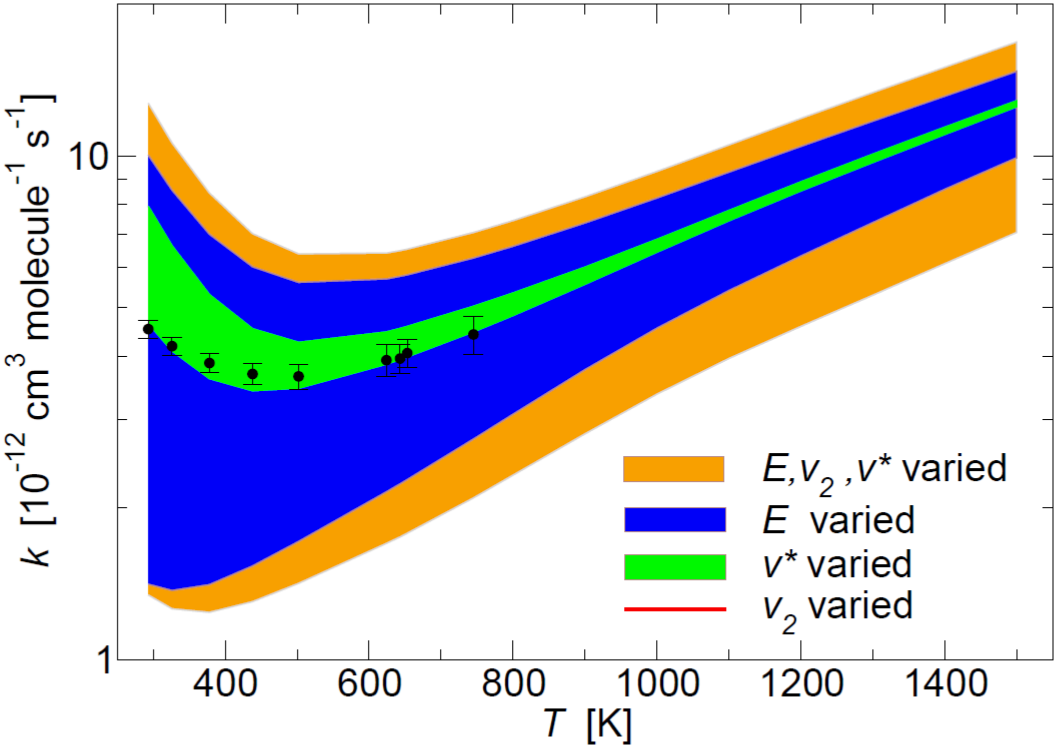
Ranges: E : -2.0...-0.7 kcal mol^{-1}, v_2 : 70...150 cm^{-1}, v^* : 50...850 cm^{-1}



- 95% intervals
- Independent, uniform distribution
- Ratio of upper to lower bound:
  (a) 10 at 300K
  (b) 2.5 at 800K
- *E* dominates up to 800K
- v_2 small temperature dependence
- v^* minor importance, only to 500K
- (R1) is challenging from theoretical/computational aspect

Uncertainty propagation using uncorrelated QC parameters:

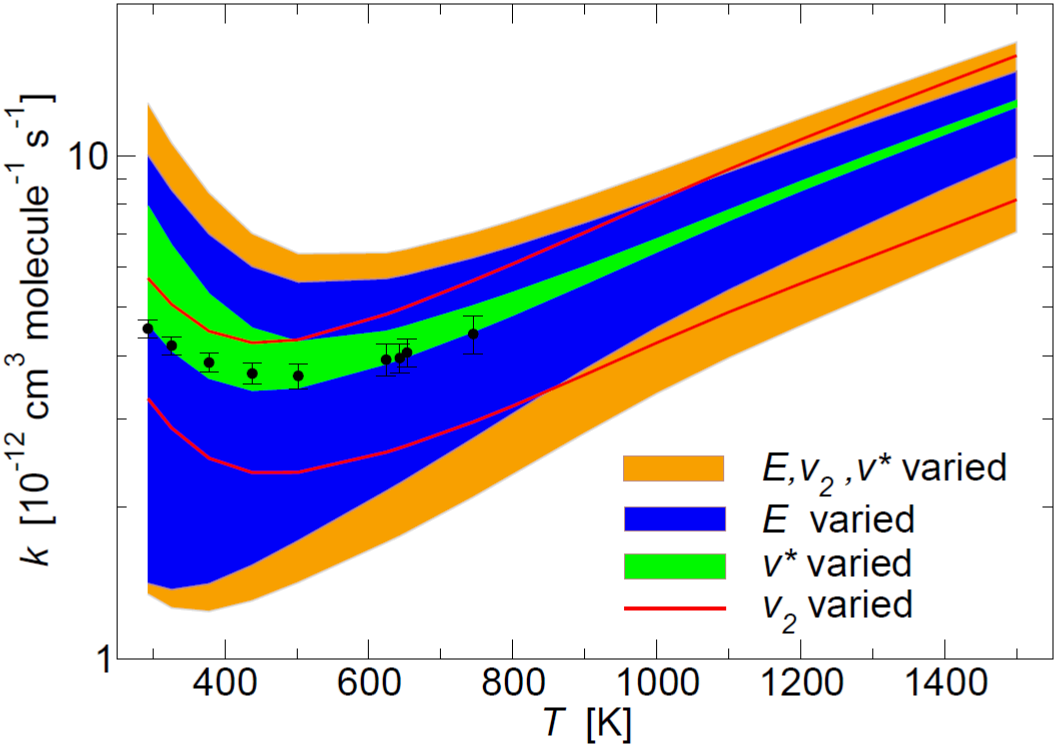
Ranges: E : -2.0...-0.7 kcal mol^{-1}, v_2 : 70...150 cm^{-1}, v^* : 50...850 cm^{-1}



- 95% intervals
- Independent, uniform distribution
- Ratio of upper to lower bound:
  - (a) 10 at 300K
  - (b) 2.5 at 800K
- *E* dominates up to 800K
- v_2 small temperature dependence
- v^* minor importance, only to 500K
- reaction is challenging from theoretical/computational aspect

Uncertainty propagation using uncorrelated QC parameters:

Ranges: *E* : -2.0...-0.7 kcal mol^{-1}, v_2 : 70...150 cm^{-1}, v^* : 50...850 cm^{-1}



- 95% intervals
- Independent, uniform distribution
- Ratio of upper to lower bound:
  - (a) 10 at 300K
  - (b) 2.5 at 800K
- *E* dominates up to 800K
- v^* minor importance, only to 500K
- v_2 small temperature dependence
- (R1) is challenging from theoretical/computational aspect

Experimental data source

Dunlop and Tully (JPC 1993) measured site-specific abstraction rate coefficients (293-745 K) and the kinetic isotope effect (KIE) using OH LIF

$$\mathrm{CH_3CH(OH)CH_3 + OH} \\ \rightarrow \mathrm{CH_3C^{\cdot}(OH)CH_3 + H_2O} \qquad \text{(R1)}$$

$$\mathrm{CD_3CD(OD)CD_3 + OH} \\ \rightarrow \mathrm{CD_3C^{\cdot}(OD)CD_3 + HDO} \qquad \text{(R2)}$$

- Very accurate and precise dataset
- Pseudo-first-order kinetic conditions
  $\rightarrow$ largest source of error:
  initial concentration of isopropanol
  $\rightarrow$ systematic error leads to shift in
  Arrhenius plot
- Error bars consider systematic errors

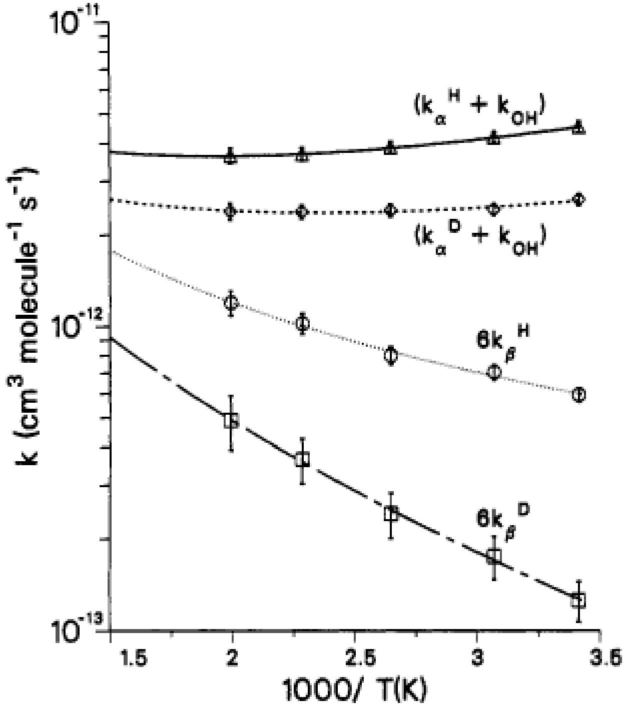


Figure 4. Site-specific rate coefficients calculated from measurements of k_a through k_e. Error bars represent $\pm 2\sigma$ estimates of the total experimental uncertainty. (—) Best fit of $\{k_\alpha^H + k_{OH}\}(T) = 2.32 \times 10^{-15} T^{1.01} \exp(538/T)$ cm^3 molecule^{-1} s^{-1}; (- - -) best fit of $\{k_\alpha^D + k_{OH}\}(T) = 1.22 \times 10^{-15} T^{1.08} \exp(450/T)$ cm^3 molecule^{-1} s^{-1}; (···) best fit of $6k_\beta^H(T) = 7.10 \times 10^{-17} T^{1.54} \exp(94/T)$ cm^3 molecule^{-1} s^{-1}; (– - –) best fit of $6k_\beta^D(T) = 5.14 \times 10^{-17} T^{1.59} \exp(-353/T)$ cm^3 molecule^{-1} s^{-1}.

Bayesian inference

- **Use experimental data** from Dunlop and Tully to obtain probabilistic description of parameters E, v_2, v^*
- Gives 'best-fit' values (MAP) and **smaller** parameter uncertainties
- Helps assess the quality of QC parameters/methods
- Can assist in model/sub-model comparisons

Uncertainty of experimental data:

$$\ln k_i(T_i) = \ln M(T_i, \Lambda) + \varepsilon_i$$

$$\varepsilon_i \sim \mathcal{N}(\alpha, \beta^2)$$

- Consider logarithm of rate coefficients
- Gaussian noise includes systematic shift α

Bayes formula:

$$p(\Lambda|D) \;\propto\; p(D|\Lambda) \cdot \pi(\Lambda)$$

- Obtain posterior distribution
- Need likelihood function and prior

Likelihood:

$$p(D|\Lambda) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\beta^2}} \times \exp\left(-\frac{(\ln k_i - \alpha - \ln M(T_i, \Lambda))^2}{2\beta^2}\right)$$

Priors:

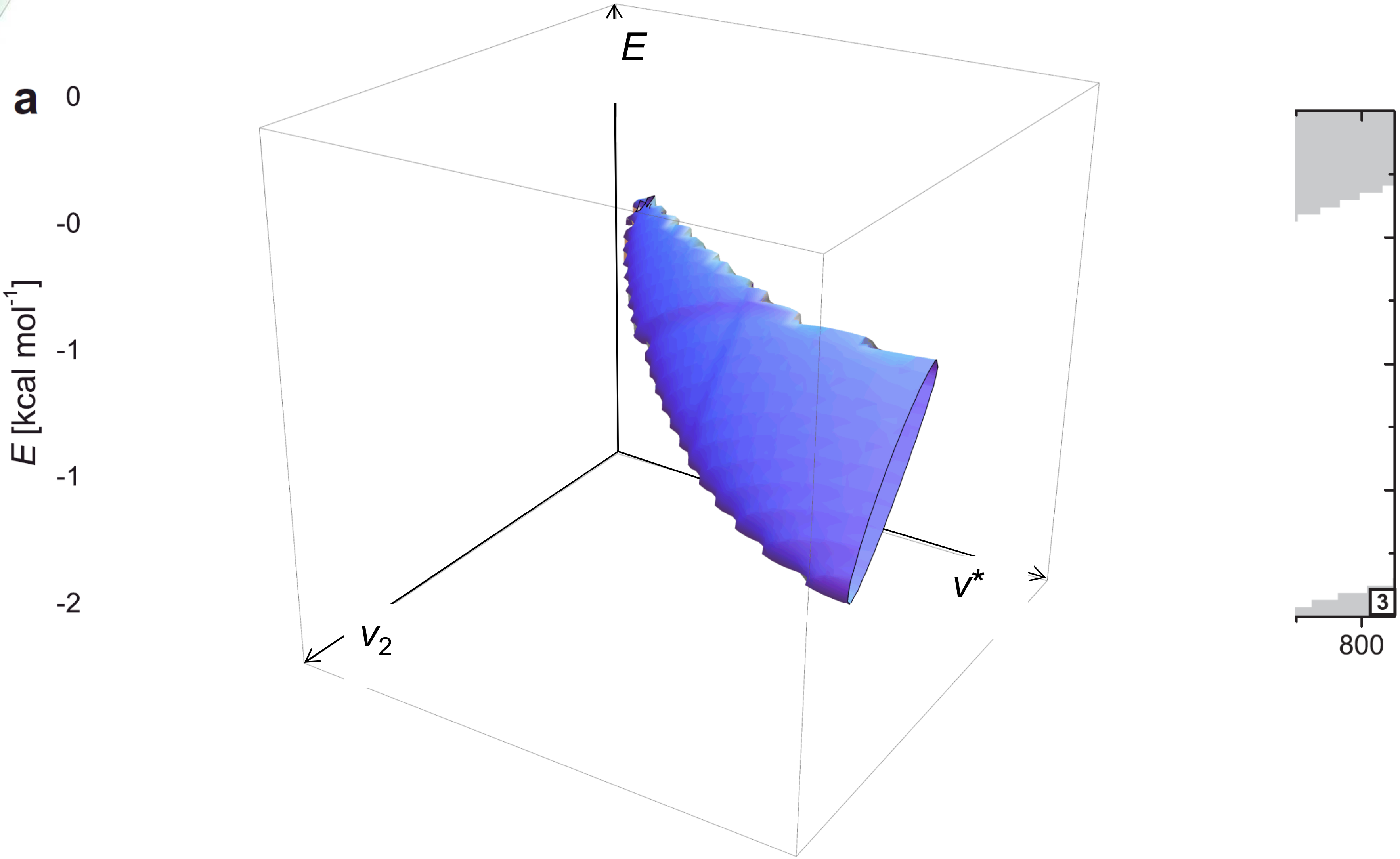
$$\pi(\Lambda, \alpha, \beta)$$

- Independent, uniform priors for E, v_2, v^*, β
- Normal distribution for α with zero mean and stdv of average error bars

Sampling from PDF(E, v_2, v^*):

- Metropolis-Hastings Markov Chain Monte Carlo (MCMC)
- 98,000 chain length was needed for convergence
- Surrogate for model (7th order Legendre polynomials) built on 10,000 points, ~±10% maximum error
- Obtain PDF, mean, and standard deviation of parameters
- MAP (maximum a posteriori) estimates give 'best fit' values
- Prediction intervals for rate coefficients with new uncertainties

Marginalized 2-D PDFs

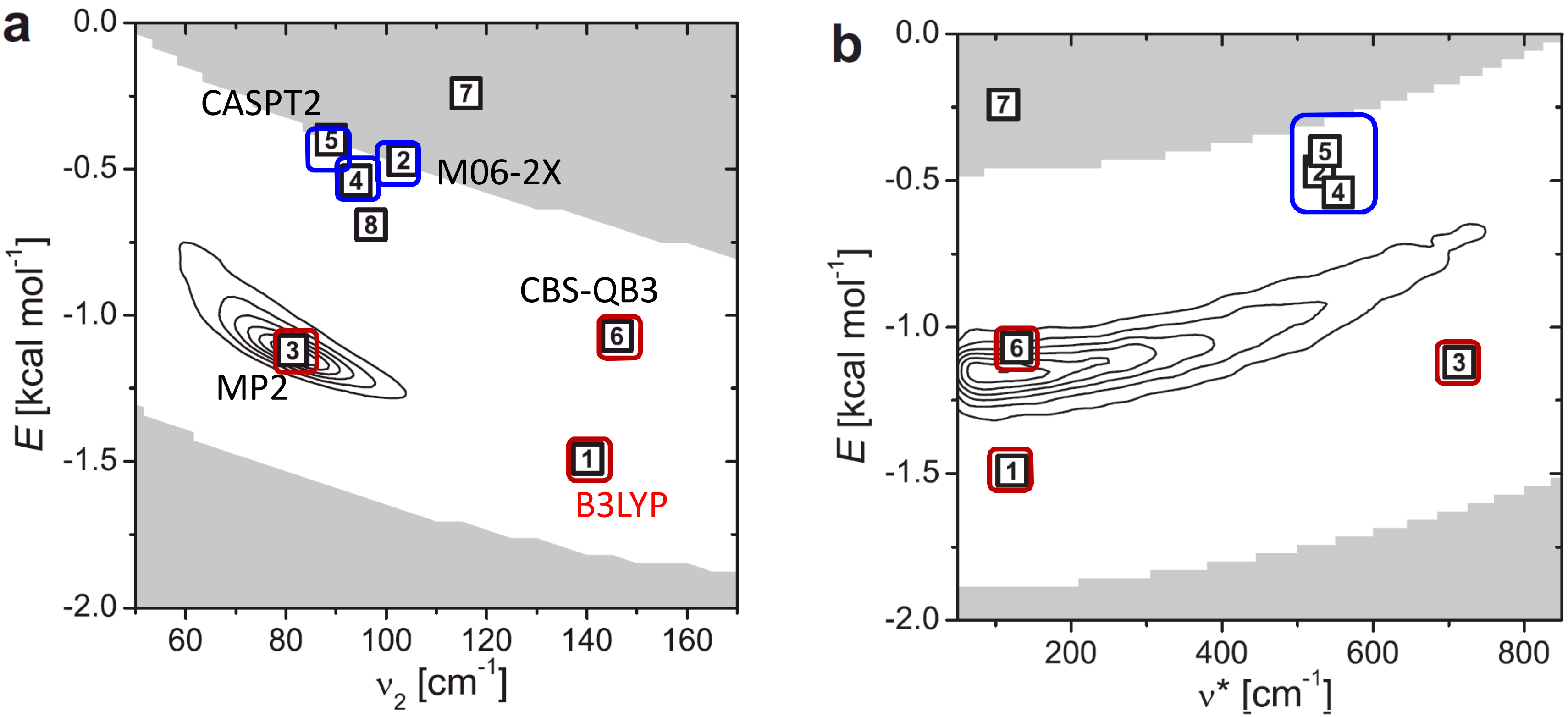


Kinetic isotope effect

IF there were no systematic discrepancies and fortuitous error cancellations, the ranking of the methods for the deuterated case should stay the same.

PDF(ν_2, E)

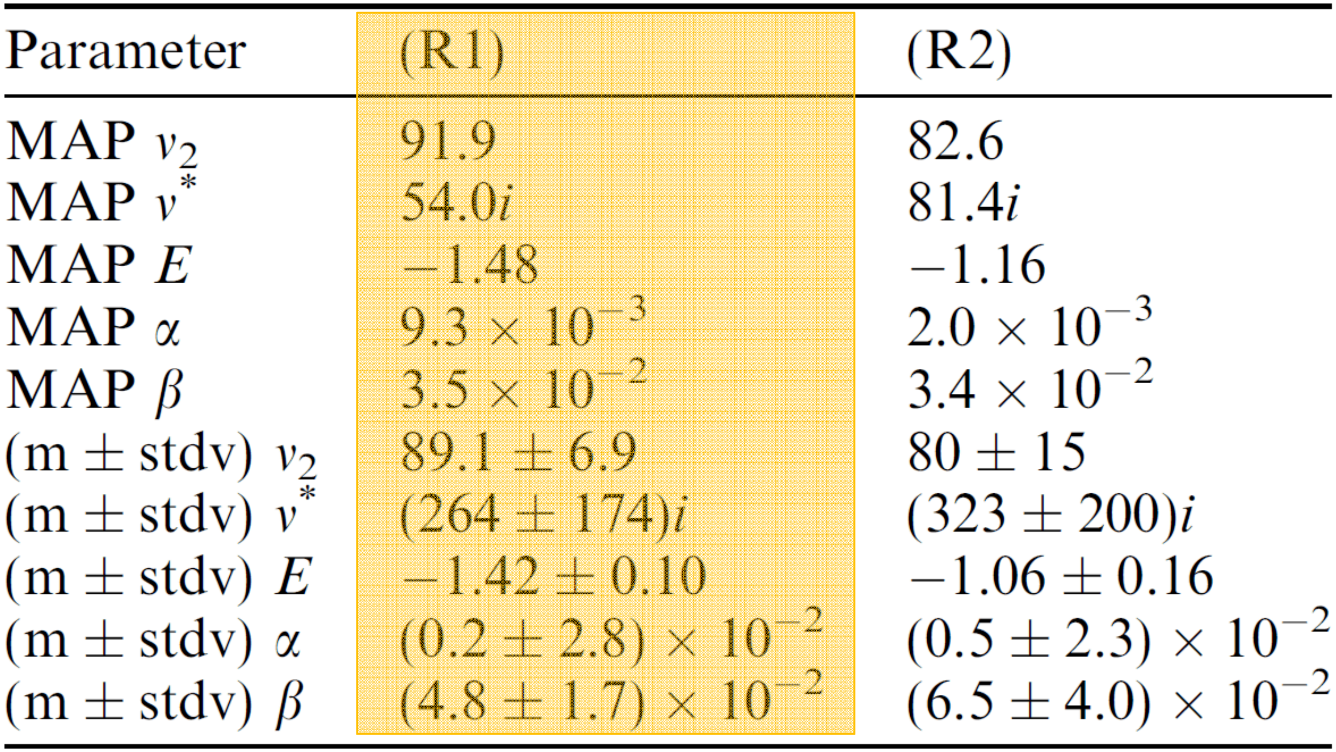
PDF(ν^*, E)



The overall picture is similar, but different methods rank better, most likely due to variational effects as well.

➔ More systematic investigation is needed to draw general conclusions.

Maximum a posteriori, mean and standard deviation estimates

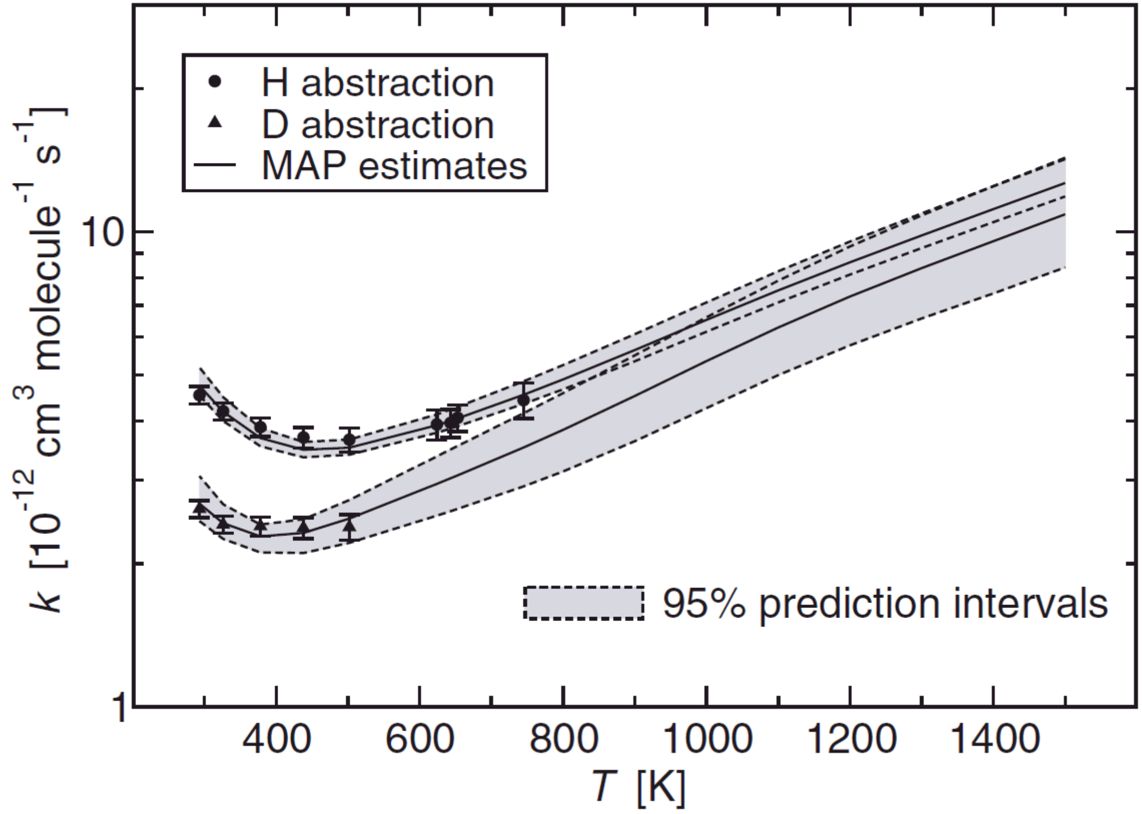


Parameter	(R1)	(R2)
MAP v_2	91.9	82.6
MAP v^*	$54.0i$	$81.4i$
MAP E	-1.48	-1.16
MAP α	9.3×10^{-3}	2.0×10^{-3}
MAP β	3.5×10^{-2}	3.4×10^{-2}
(m $\pm$ stdv) v_2	89.1 ± 6.9	80 ± 15
(m $\pm$ stdv) v^*	$(264 \pm 174)i$	$(323 \pm 200)i$
(m $\pm$ stdv) E	-1.42 ± 0.10	-1.06 ± 0.16
(m $\pm$ stdv) α	$(0.2 \pm 2.8) \times 10^{-2}$	$(0.5 \pm 2.3) \times 10^{-2}$
(m $\pm$ stdv) β	$(4.8 \pm 1.7) \times 10^{-2}$	$(6.5 \pm 4.0) \times 10^{-2}$

- Again: well defined parameters *E* and v_2 , uncertain v^*
- The stdv of α (mean of the shift in the data) did not decrease during the inference.

Predicted uncertainty in extrapolated calculated rate coefficients is obtained

Temperature dependent *f*-factors for numerical simulations are conveniently obtained for theoretically extrapolated rate coefficients!



- Prediction bands are consistent with the data and error bars
- Updated uncertainties are greatly reduced
- TST model is consistent with the data

Alkane + OH series

- Isopropanol study was motivated by our work on alcohols, but alcohol + OH reactions are more complicated due to coupling of OH groups
- We do not limit ourselves to 3 parameters
- Correlations will be explored better, both in forward uncertainty and in Bayesian priors
- There is more data available on alkanes, therefore, are better candidates for systematic investigations

ethane + OH

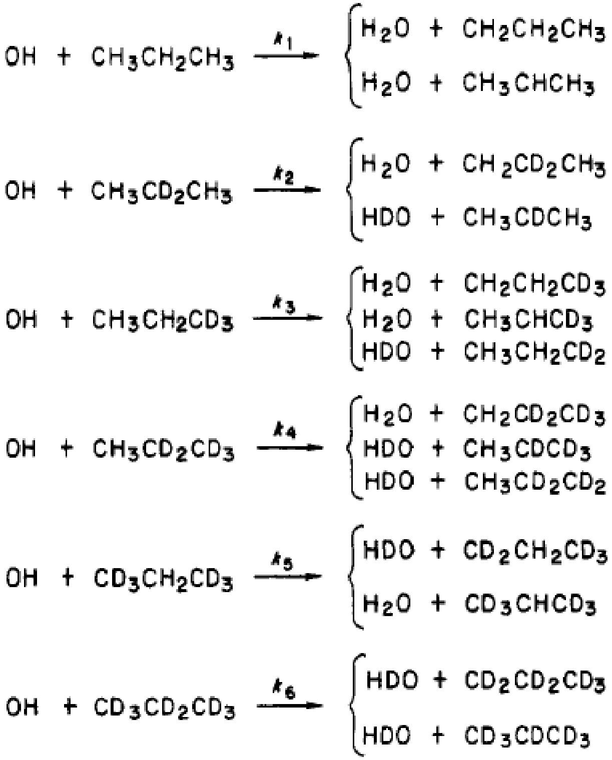
$C_2H_6 + OH = C_2H_5 + H_2O$

$C_2D_6 + OH = C_2D_5 + DHO$

$CH_3CD_3 + OH = CH_2CD_3 + H_2O$

$CH_3CD_3 + OH = CH_3CD_2 + DHO$

propane + OH

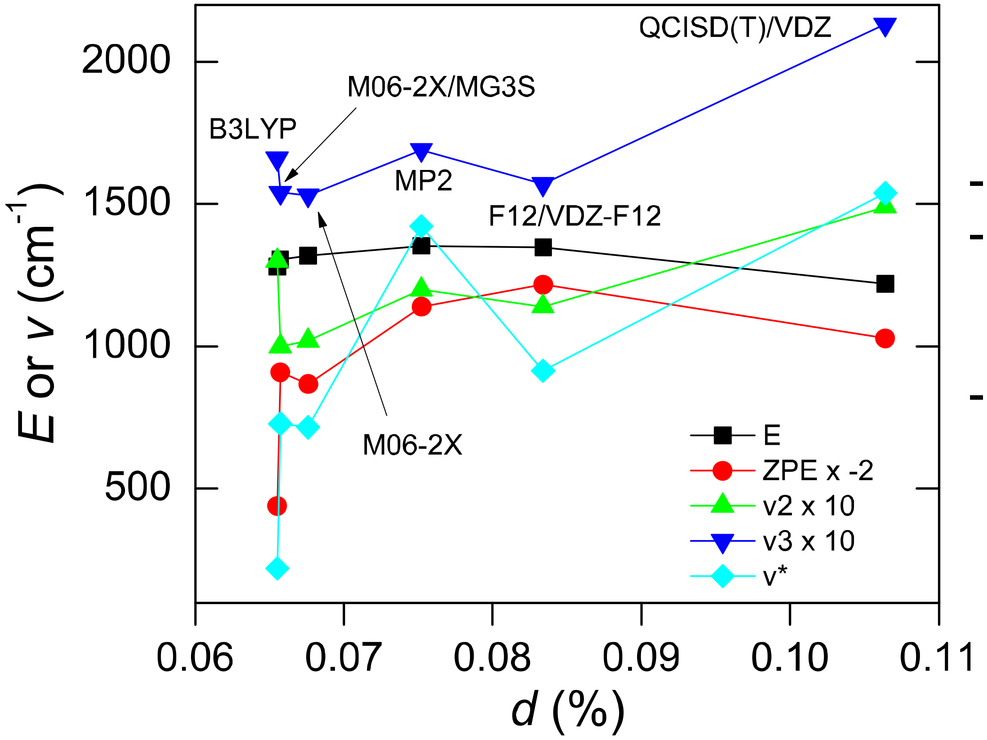


butane + OH

isotopic labeling and structural isomers…

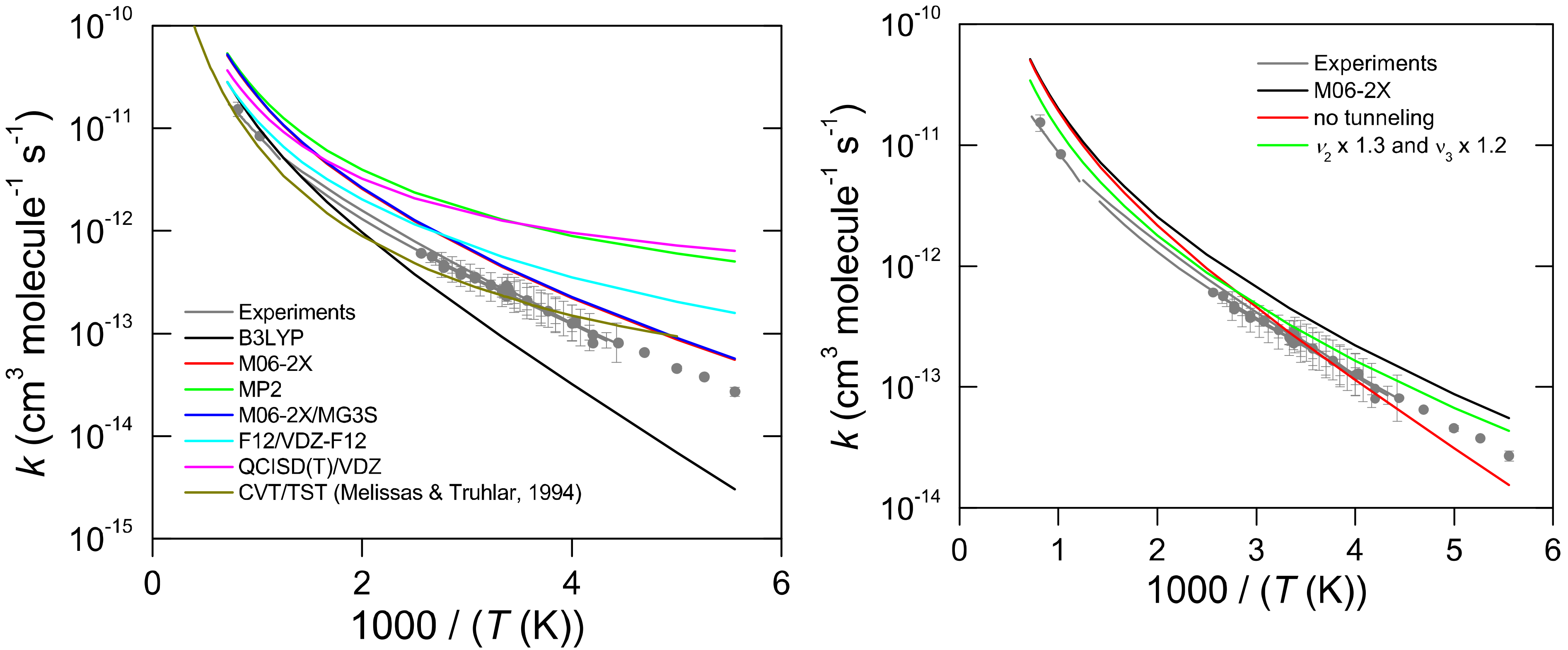
Ethane + OH

	QCISD(T)/ CBS	UCCSD(T)-F12/ cc-pVTZ-F12	ZPE(H)	A	B	C	d%	v*	v1	v2
B3LYP/6-311++G(d,p)	3.66	3.64	-0.63	26.8	4.6	4.1	7%	-220	50	130
M06-2X/6-311++G(d,p)	3.77	3.77	-1.30	25.8	4.8	4.3	7%	-716	73	102
MP2/6-311++G(d,p)	3.87		-1.63	26.3	4.9	4.4	8%	-1421	79	120
M06-2X/MG3S	3.73	3.73	-1.24	26.1	4.8	4.3	7%	-729	42	100
UCCSD(T)-F12a/cc-pVDZ-F12	3.85		-1.74	26.2	4.9	4.3	8%	-915	48	114
QCISD(T)/pVDZ	3.49		-1.47	24.6	5.1	4.6	11%	-1538	113	149



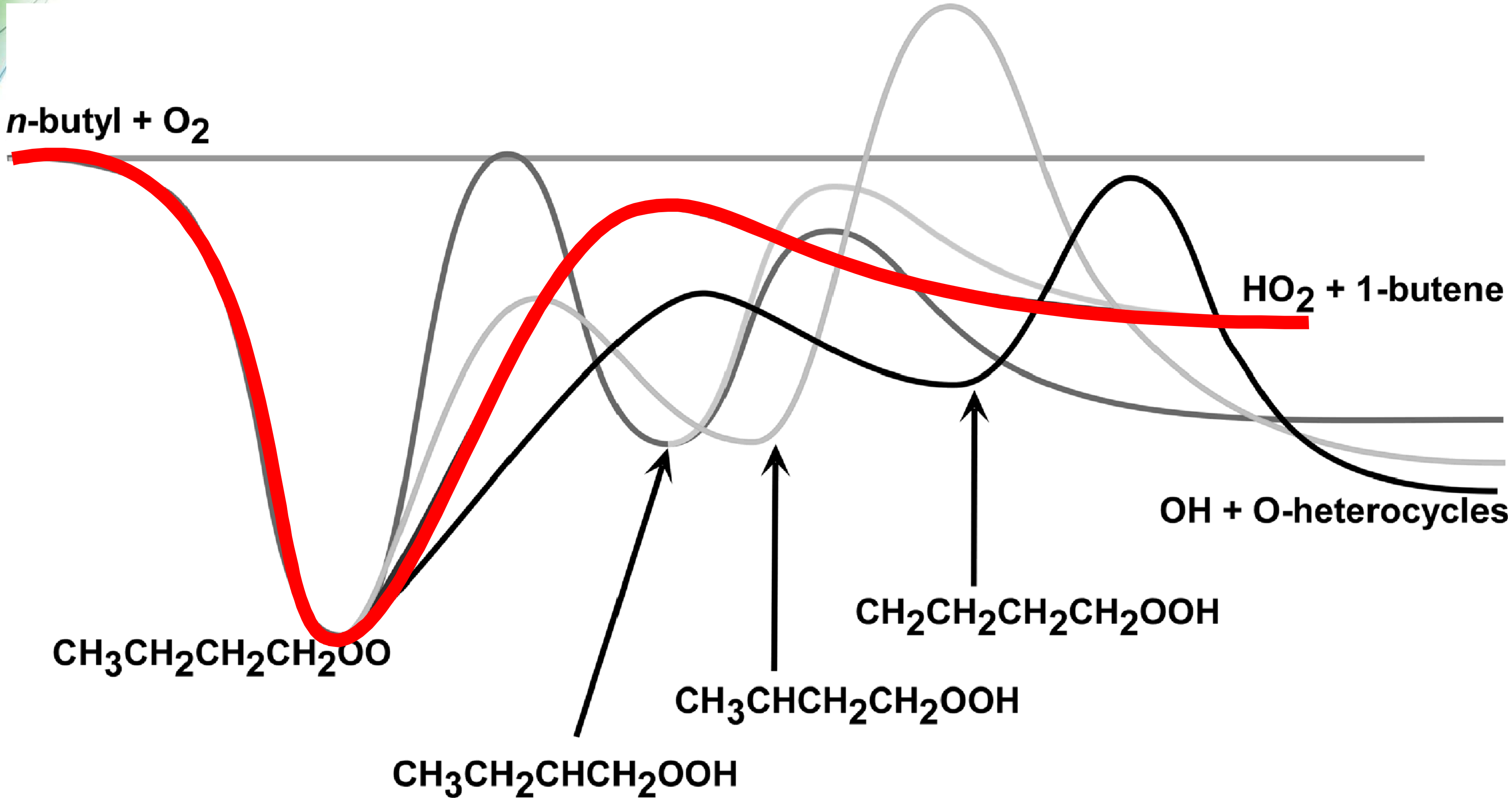
- Correlation is less clear
- Most of the apparent uncertainty arises due to uncertainties in frequencies
- Electronic energies are close to CBS, but 0.5-1 kcal/mol systematic uncertainty is still likely

Preliminary forward uncertainty estimates of the calculated ethane + OH rate coefficients



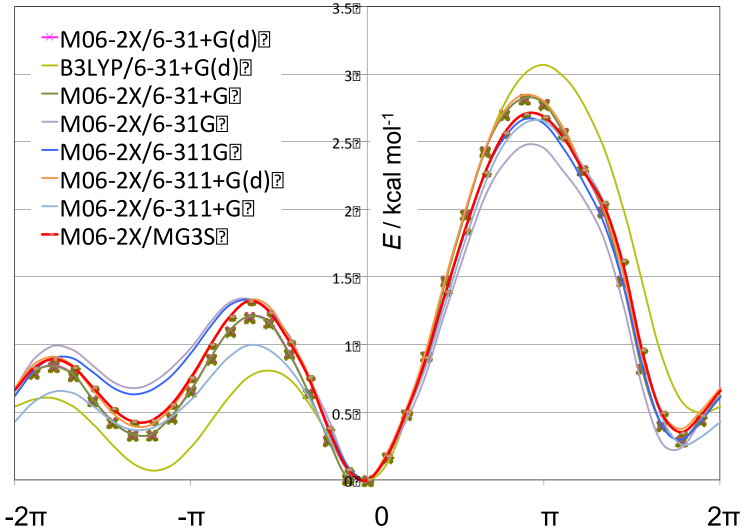
- For this reaction M06-2X clearly performs the best, but it remains to be seen whether this is true for the whole set
- Tunneling model is critical

Further uncertainties arising in larger systems: the *n*-butyl + O_2 reaction

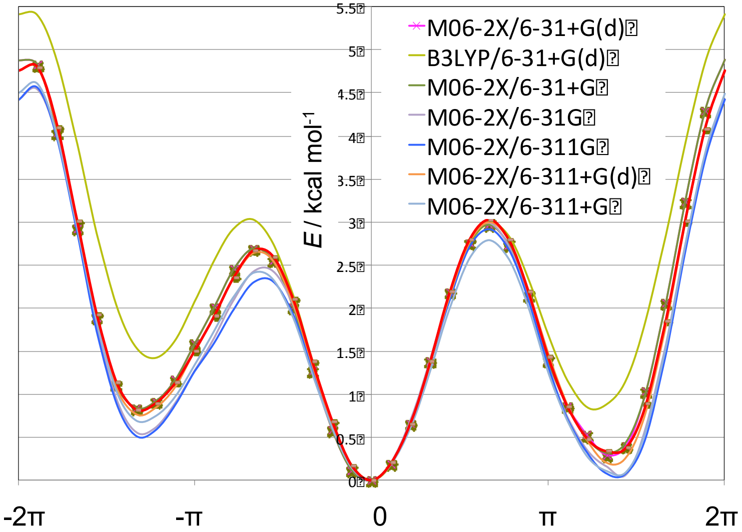


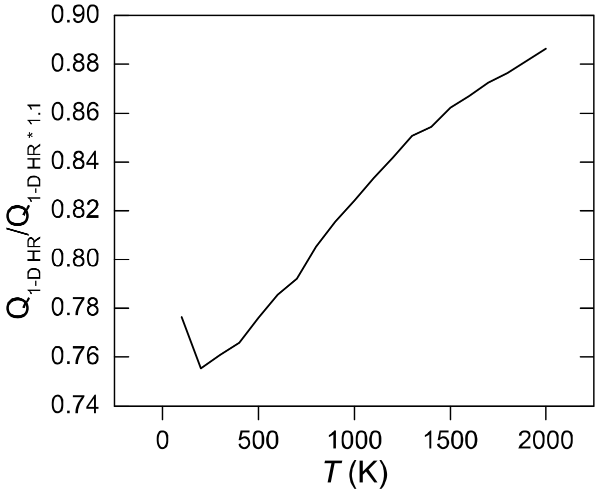
1-D separable hindered rotor approximation

O-O-C-C

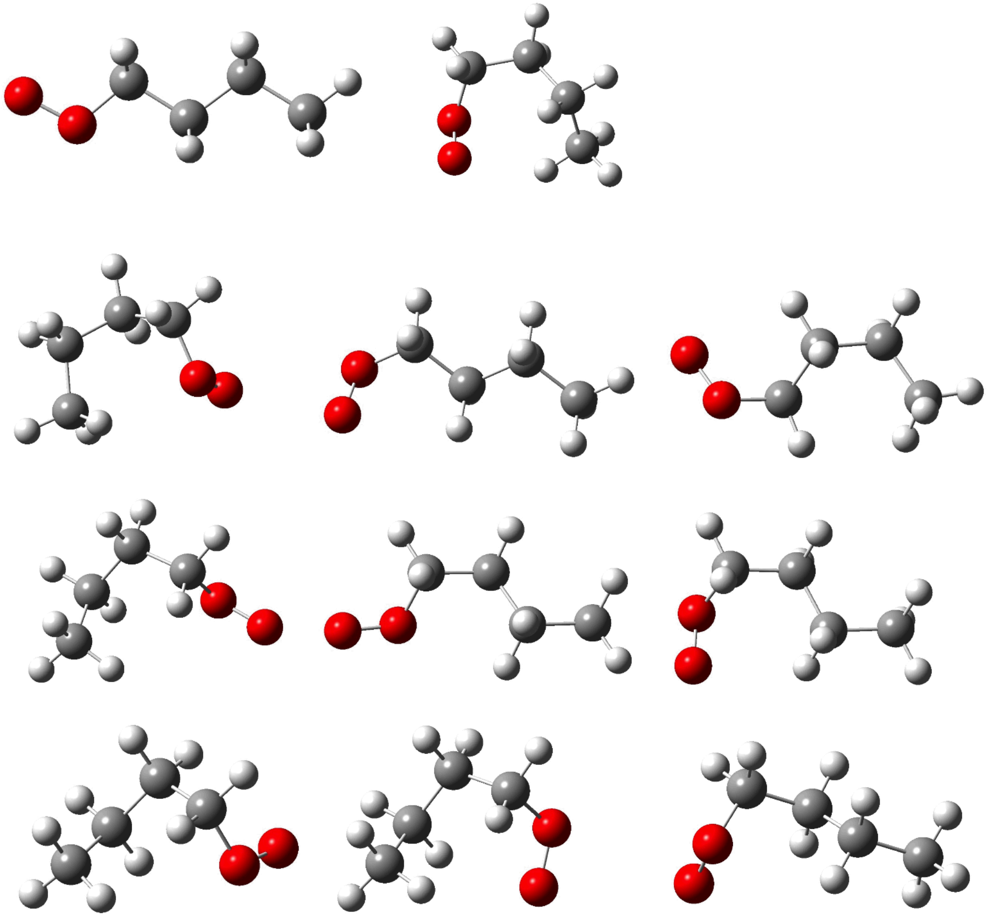


C-C-C-C

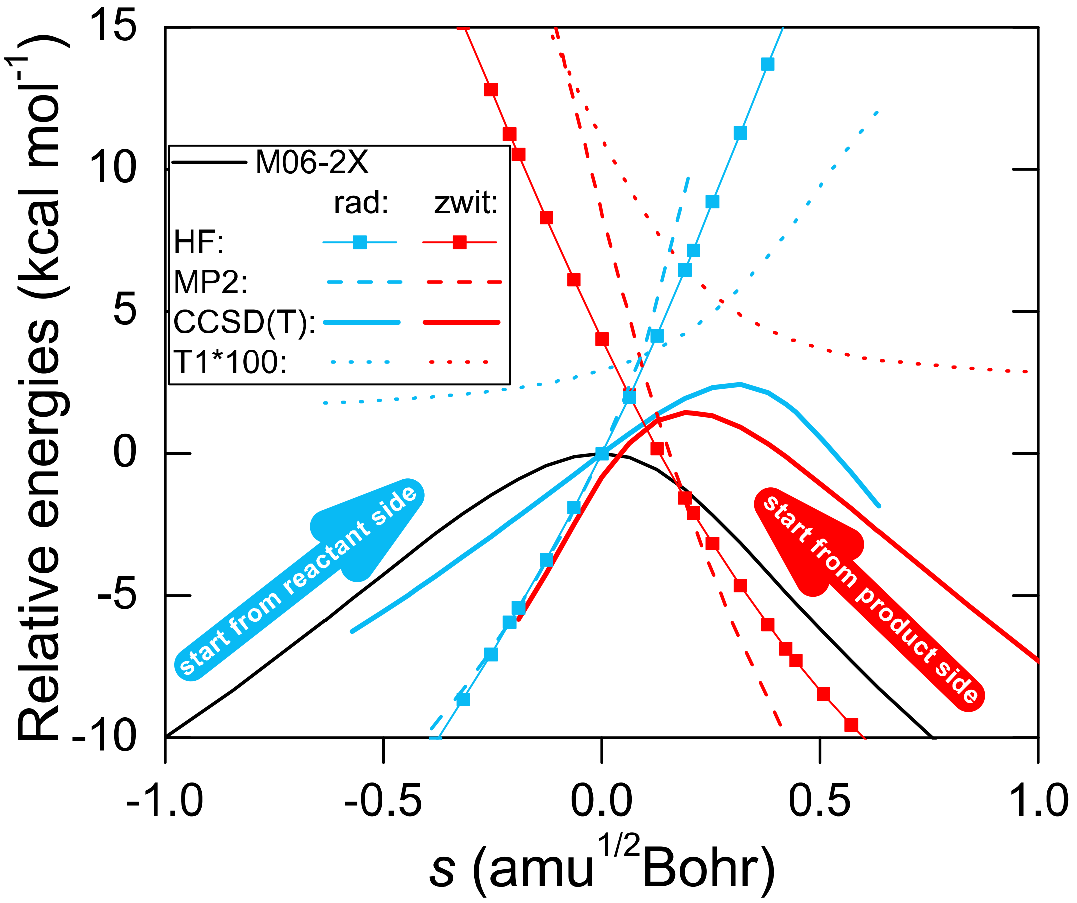




Thermal rate coefficients are highly averaged macroscopic quantities, both fundamentally and pragmatically



In the single reference framework two weakly interacting states can be found



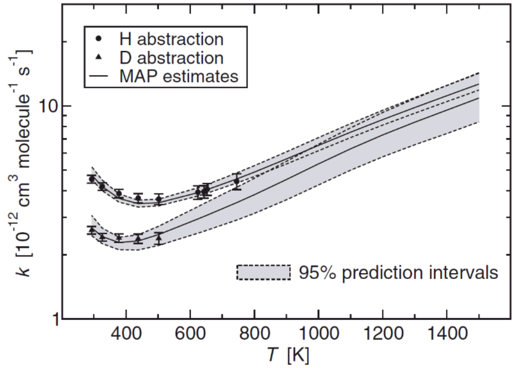
reactant side: **radical** character, small dipole moment
product side: **zwitterion** character, larger dipole moment

dipole moments (D)

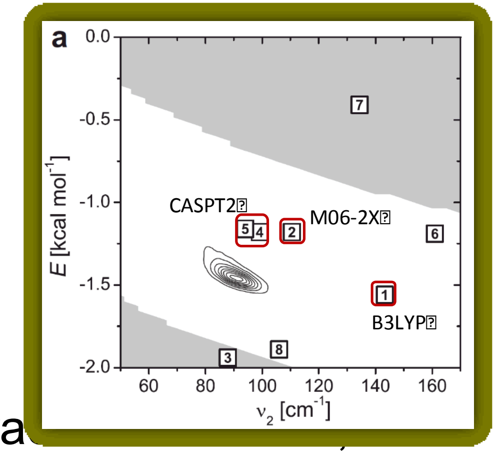
	radical	**zwitterion**
ROHF:	2.1	4.9
RMP2:	1.3	4.9
UCCSD:	1.1	2.9
UCCSD(T):	1.4	1.6

UCCSD(T) energies become less dependent on initial orbital guesses
barriers for ethanol case: 18.2 and 17.2 kcal mol^{-1}
M06-2X barrier: 16.9 kcal mol^{-1} ➔ close (within uncertainty) to the UCCSD(T) barriers.

Conclusions and outlook



- **Bayesian inference**, when used in connection to reliable and well-characterized data, is a very **versatile** and **rigorous tool** to
  - uncover weak parts of theoretical kinetic models
  - extrapolate both rate coefficients and uncertainty bounds (*T*-dependent *f*-factors)



- Barrier heights are not any more the single largest sources of uncertainty in theoretical kinetics calculations.
- Our methodology will be further refined by
  - pre-screening parameters to be selected for inference
  - accommodate correlations among parameters
  - enabling direct sampling (no need for surrogate)
- Starting with simple, experimentally well-characterized abstra develop a **benchmark kinetic data set** of reactions and the corresponding **computational tools** to systematically and **consistently** develop our theoretical approaches, and **go beyond factor of two uncertainty**.

Conclusions and outlook

- We need to go beyond MC uncertainty studies, as the fundamental parameters feeding into the calculations are correlated.
- Complicated electronic structure cases are equivalents of missing reactions
- conformers – how to treat them

Conclusions and outlook

- electronic energy ~ large multireference character ~ pathological transition states
- ZPE ~ high frequencies ~ location of the TS
- state counts ~ anharmonicities ~ coupled torsions ~ conformers
- variational treatment
- tunneling (not so relevant for combustion)
- collisional model

Thank You!

and thank our sponsors:

Division of Chemical Sciences, Geosciences, and Biosciences, the Office of Basic Energy Sciences, the U.S. Department of Energy under DOE Contract Numbers DE-AC04-94AL85000.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the National Nuclear Security Administration.