

# xtb meets ONIOM & Iterative solver for eigenvalue problem

Albert Katbashev

Mulliken Center for Theoretical Chemistry  
University of Bonn

May 17, 2023



# Outline

**1** Motivation

**2** Theory

**3** Best practice

**4** Results

**5** Summary and outlook



# Motivation

Size	FF		SQM		DFT	
	acc.	eff.	acc.	eff.	acc.	eff.
1-300	✗	✓	~	✓	✓	✓
300-3000	✗	✓	~	✓	✓	✗
>3000	✗	✓	~	~	✓	✗



# Motivation

Size	FF		SQM		DFT	
	acc.	eff.	acc.	eff.	acc.	eff.
1-300	✗	✓	~	✓	✓	✓
300-3000	✗	✓	~	✓	✓	✗
>3000	✗	✓	~	~	✓	✗

For large molecules:

- SQM, FF feasible, but not accurate enough
- DFT accurate, but not feasible

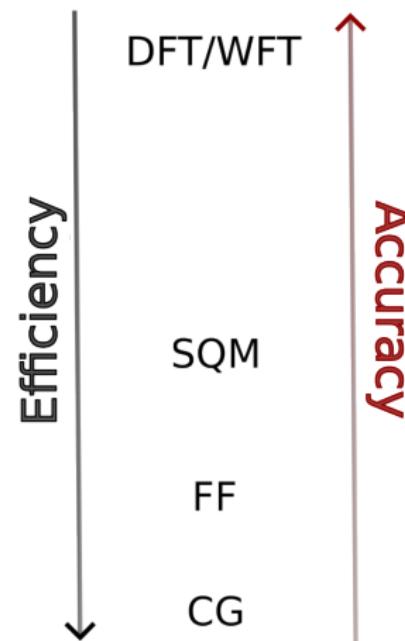


# Motivation

Size	FF		SQM		DFT	
	acc.	eff.	acc.	eff.	acc.	eff.
1-300	✗	✓	~	✓	✓	✓
300-3000	✗	✓	~	✓	✓	✗
>3000	✗	✓	~	~	✓	✗

For large molecules:

- SQM, FF feasible, but not accurate enough
- DFT accurate, but not feasible



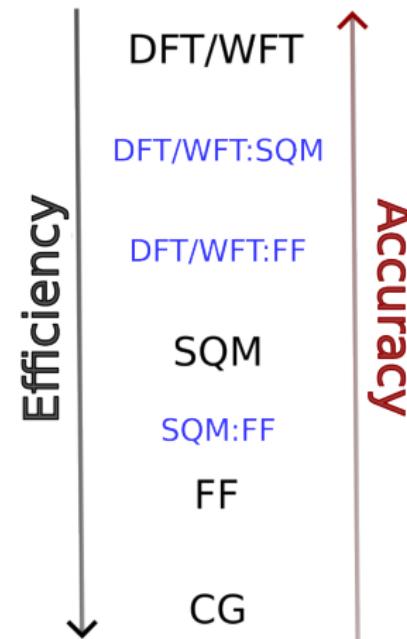


# Motivation

Size	FF		SQM		DFT	
	acc.	eff.	acc.	eff.	acc.	eff.
1-300	✗	✓	~	✓	✓	✓
300-3000	✗	✓	~	✓	✓	✗
>3000	✗	✓	~	~	✓	✗

For large molecules:

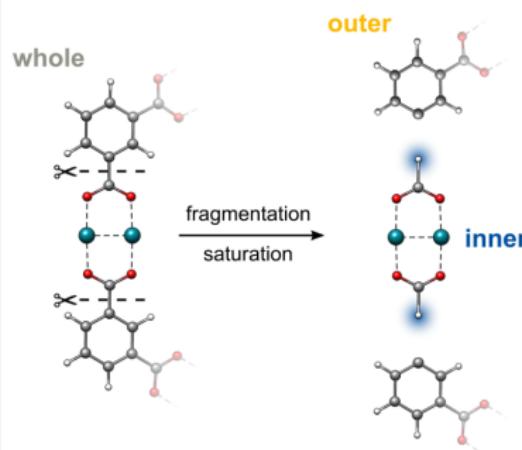
- SQM, FF feasible, but not accurate enough
- DFT accurate, but not feasible
- Possible solutions:  
QM/MM, ONIOM, Subsystem DFT





# 2-Layer ONIOM/Link Atom approach

Our own N-layered Integrated  
molecular Orbital and Molecular mechanics



$$E_{\text{ONIOM}} = E_{\text{inner}}^{\text{high}} + E_{\text{whole}}^{\text{low}} - E_{\text{inner}}^{\text{low}}$$



# Application guideline

- 1 Suitability check: *high-level – inner region, low-level – whole system*
- 2 Implicit solvation only for *low-level – whole system*
- 3 Test-run: check geometry and charges (`--cut`)
- 4 Boundary handling -> only  $\sigma$  bond cleavage
- 5 The input for the external software has to be scrutinized

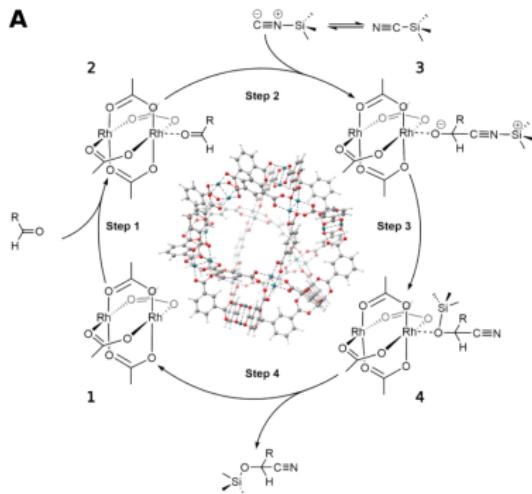
## usage

```
xtb <geometry_file> --onionm high:low <inner_region>
```



# Electronic Energy

System: DALTES

**A**

Size: 408+22



Motivation  
□□□□

Theory  
□□

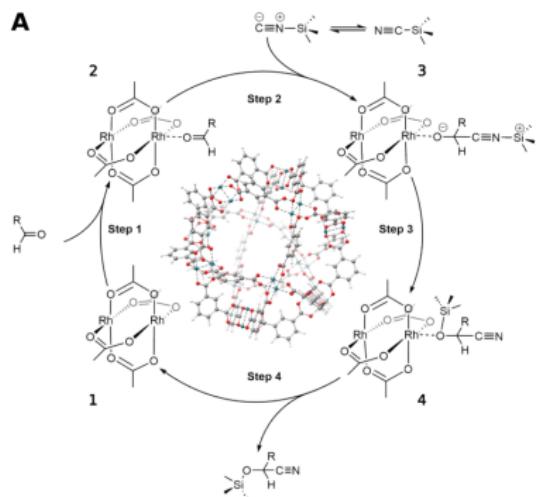
Best practice  
□□

Results  
□□□□□

Summary and outlook  
□

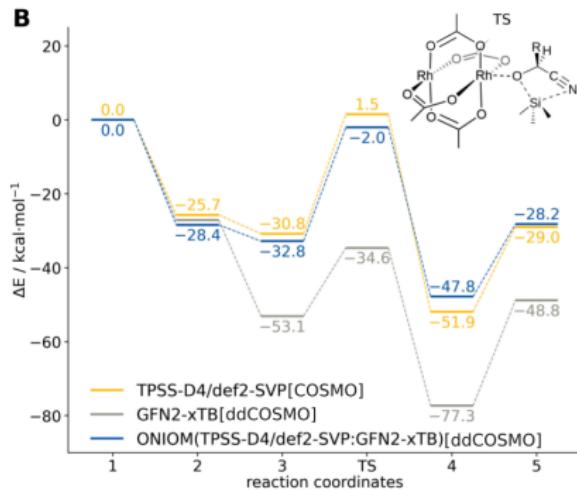
# Electronic Energy

System: DALTES



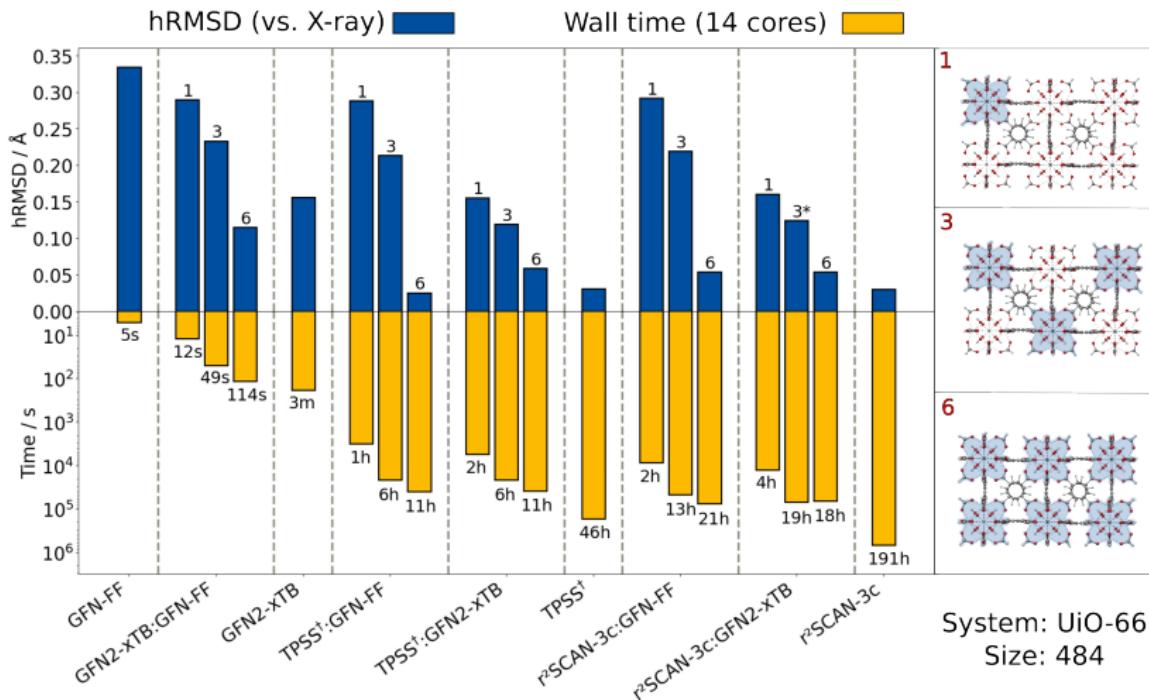
Size: 408+22

Reaction: cyanosilylation



Geometries: TPSS-D4/def2-SVP

## Geometry Optimization

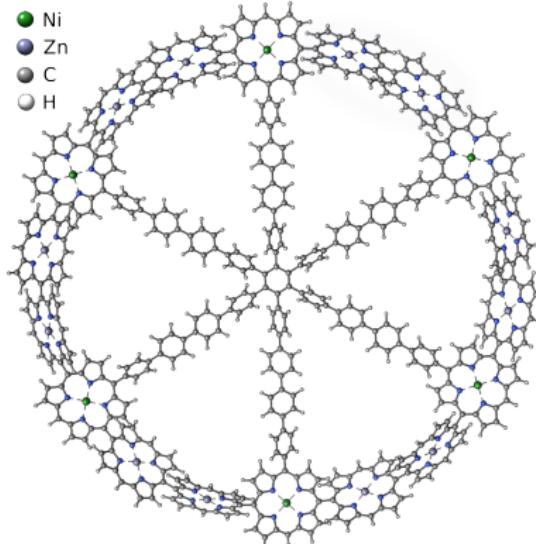




# Geometry Optimization + Electronic Energy

System: Spoked porphyrin ring

- Ni
- Zn
- C
- H



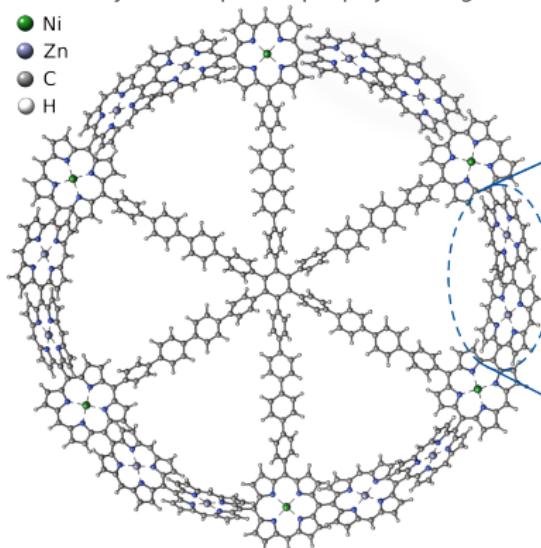
Size: 870



# Geometry Optimization + Electronic Energy

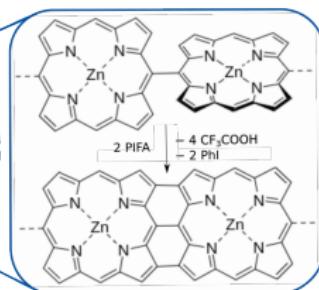
System: Spoked porphyrin ring

- Ni
- Zn
- C
- H



Size: 870

C-C coupling

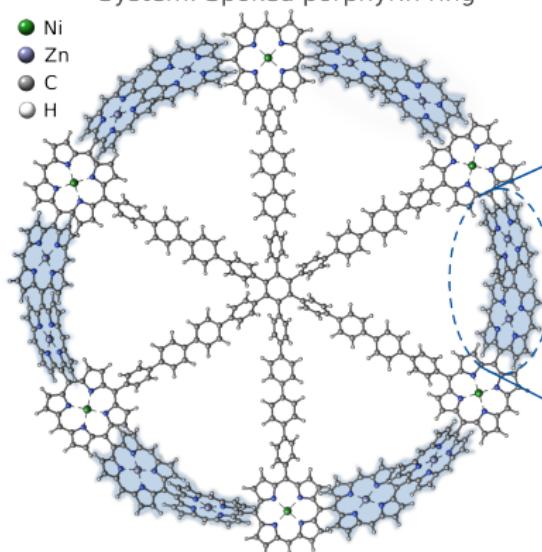




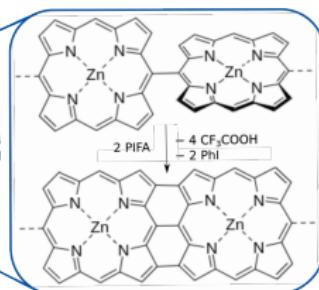
# Geometry Optimization + Electronic Energy

System: Spoked porphyrin ring

- Ni
- Zn
- C
- H



C-C coupling



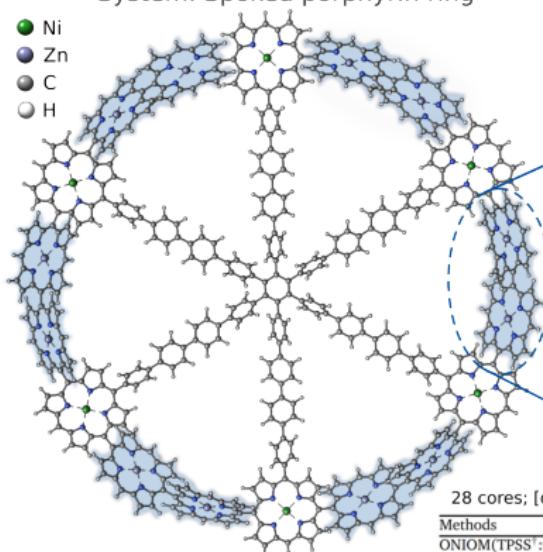
Size: 870  
(Inner region 472 )



# Geometry Optimization + Electronic Energy

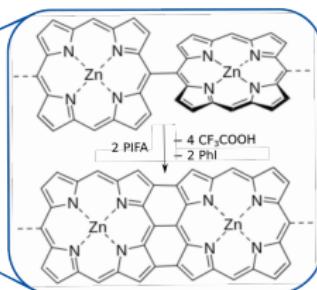
System: Spoked porphyrin ring

- Ni
- Zn
- C
- H



Size: 870  
(Inner region 472 )

C-C coupling



28 cores; [ddCOSMO];  $\text{TPSS}^\dagger = \text{TPSS-D4}/\text{def2-SVP}$

Methods	$\Delta E / \text{kcal}\cdot\text{mol}^{-1}$	SCF time
ONIOM(TPSS <sup>†</sup> :GFN2-xTB)	-44.9	25m
GFN2-xTB	-34.9	36s
ONIOM(GFN2-xTB:GFN-FF)	-24.4	4s
ONIOM(TPSS <sup>†</sup> :GFN1-xTB)	-47.0	24m
GFN1-xTB	-37.6	39s
ONIOM(GFN1-xTB:GFN-FF)	-26.6	4s
ONIOM(TPSS <sup>†</sup> :GFN-FF)	-35.3	15m
GFN-FF	172.6	0.4s



# Summary and outlook

- ✓ ONIOM implementation in xtb: single-point energy, geometry, frequency
- ✓ Extensively tested for metal-organic moieties
- ✓ Paper (in revision)

## TO DO:

- ✗ Beyond 2-layers
- ✗ Embedding
- ✗ Proteins, nucleic acids, carbohydrates



## Iterative solver for eigenvalue problem



# Outline

**6** Motivation

**7** Theory

**8** Plan

**9** Summary and outlook



# Motivation

## SCF

$FC = SC\epsilon$  (general)  
 $F'C' = C'\epsilon$  (special)

Density matrix  
purification

Diagonalization  
`dsyev`  
complexity:  $O(N^{2-2.5})$   
limited parallelism

Matrix multiplication  
`dgemm`  
complexity:  $O(N^3)$   
high parallel efficiency

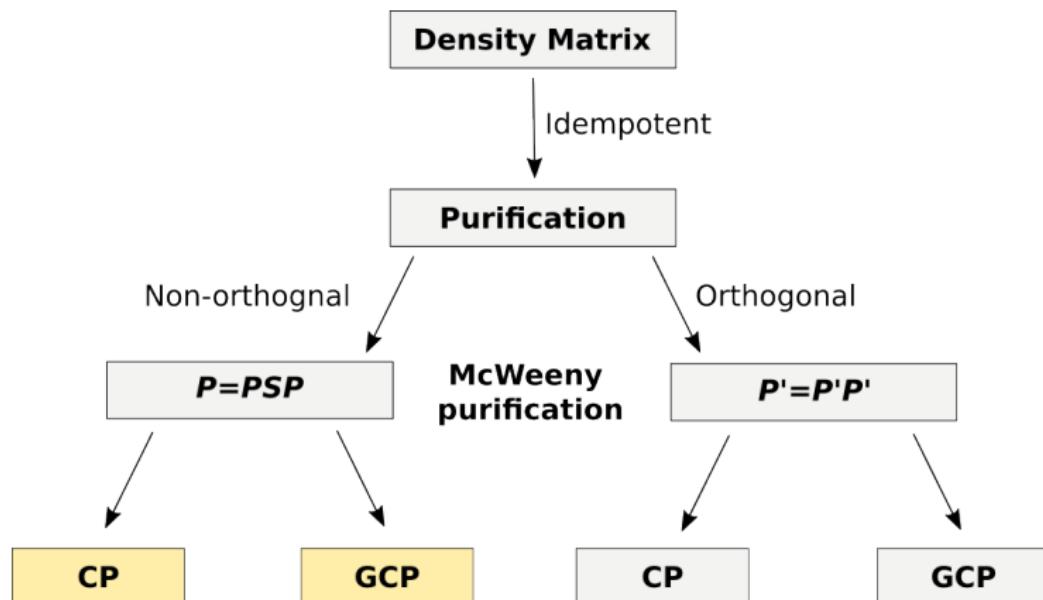
small & medium

large

[1]



# McWeeny Purification



**CP:** Canonical Purification

**GCP:** Grand Canonical Purification



# Ensembles

## Grand Canonical Purification

$$p_0 = \frac{1}{2}(\alpha(\mu S^{-1} - S^{-1}HS^{-1}) + S^{-1})$$

$$p_{k+1} = 3p_kSp_k - 2p_kSp_kSp_k$$

$\mu = \text{const}$

$k = 0, 1, 2, \dots$



# Ensembles

## Grand Canonical Purification

$$p_0 = \frac{1}{2}(\alpha(\mu S^{-1} - S^{-1}HS^{-1}) + S^{-1})$$

$$p_{k+1} = 3p_kSp_k - 2p_kSp_kSp_k$$

$\mu = \text{const}$   $\longrightarrow$  guess

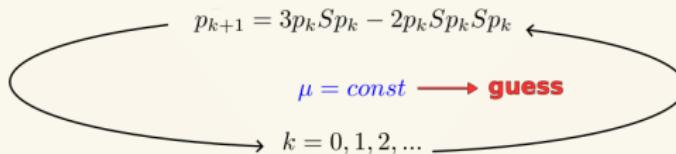
$k = 0, 1, 2, \dots$



# Ensembles

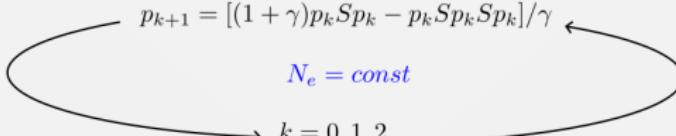
## Grand Canonical Purification

$$p_0 = \frac{1}{2}(\alpha(\mu S^{-1} - S^{-1}HS^{-1}) + S^{-1})$$

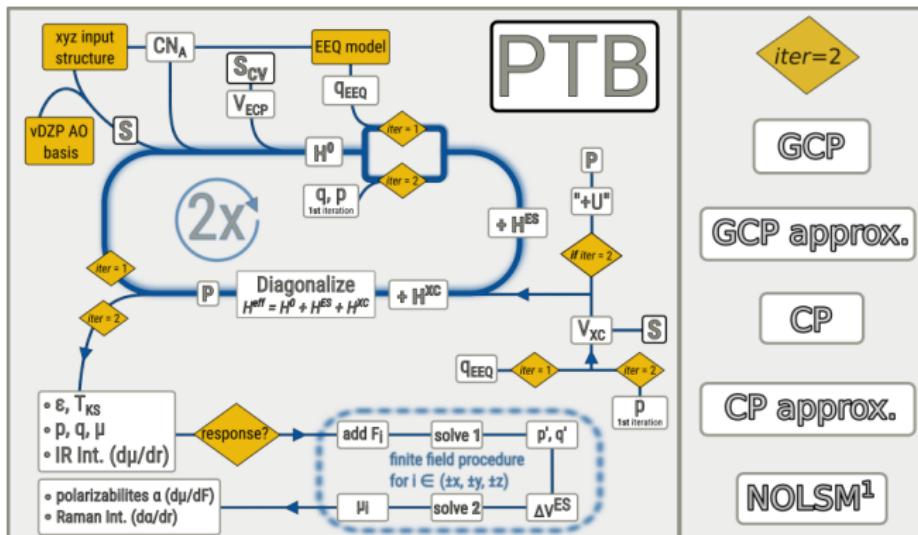


## Canonical Purification

$$p_0 = \frac{1}{N}(\alpha\mu - \alpha S^{-1}H + N_e I)S^{-1}$$



PTB



### <sup>1</sup>non-orthogonalized local submatrix method



# Summary and outlook

- Correlation between P(iterative) and P(diagonalization)
- Test on GPUs (maybe FPGAs)
- integration into xtb
- GP3



Special thanks to:

- Stefan Grimme
- Christoph Plett
- Markus Bursch
- Andreas Hansen
- Thomas Kühne
- Marcel Müller

Thank you for your attention!

# Implementation

Table: Element-specific parameters used to calculate the static scaling factor  $k$  in the ONIOM implementation of *xtb*. The average bond distances  $d^{\text{ONIOM}}$  are given in Å.

Elements	H	C	O	N	P	S
H	0.740	1.084	0.964	1.024	1.414	1.389
C	-	1.528	1.430	1.475	1.860	1.750
O	-	-	1.450	1.360	1.750	1.500
N	-	-	-	1.470	1.770	1.650

$$\vec{R}_{\text{LA}} = \vec{R}_{\text{inner}} + k(\vec{R}_{\text{outer}} - \vec{R}_{\text{inner}}) \quad (1)$$

Herein, k factor is defined as:

$$k = \frac{d_1}{d_2} \quad (2)$$

# Implementation

$$\frac{\delta R_{\text{LA}, i}}{\delta R_{\text{inner}, i}} = (1 - k), \text{ with } i=x,y,z \quad (3)$$

$$\frac{\delta R_{\text{LA}, i}}{\delta R_{\text{outer}, i}} = k \quad (4)$$

$$d_2 = \sqrt{\sum_{i=1}^3 (R_{\text{outer}, i} - R_{\text{inner}, i})^2}. \quad (5)$$

$$\frac{\delta R_{\text{LA}, i}}{\delta R_{\text{inner}, i}} = 1 - k \left( 1 - \frac{(R_{\text{outer}, i} - R_{\text{inner}, i})^2}{d_2} \right), \quad (6)$$

$$\frac{\delta R_{\text{LA}, i}}{\delta R_{\text{outer}, i}} = k \frac{\sum_{j \neq i}^3 (R_{\text{outer}, j} - R_{\text{inner}, j})^2}{d_2}. \quad (7)$$

# DALTES/optimization

Table: hRMSD (vs.X-Ray) and wall time values (on 28 cores) of the geometry optimizations of the DALTES polyhedron with different QC methods, ordered according to the increasing computing time.

Methods	hRMSD	Wall time
GFN-FF	0.315	3s
GFN2-xTB	0.178	2m
<b>TPSS-D4/def2-SVP</b>	<b>0.143</b>	<b>15h 42m</b>
TPSS-D4/def2-mSVP	0.140	37h 18m
PBEh-3c	0.167	51h 10m
B97-3c	0.162	73h 55m
r <sup>2</sup> SCAN-3c	0.152	131h 40m
TPSS-D4/def2-TZVP	0.153	147h 45m

# DALTES/r<sup>2</sup>SCAN-3c

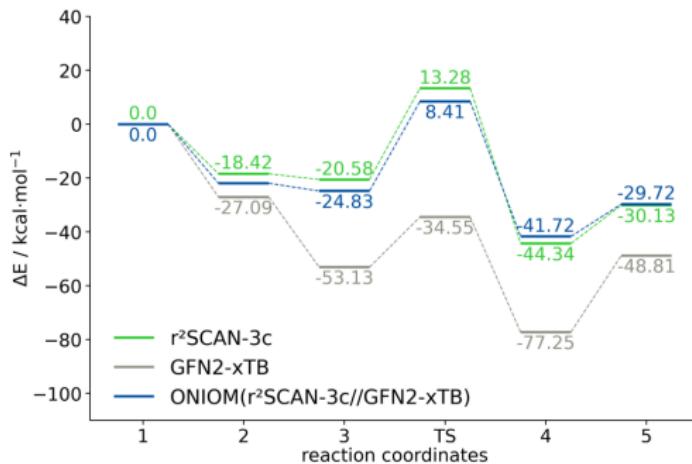


Figure: The relative potential energy curve of the cyanosilylation reaction computed with r<sup>2</sup>SCAN-3c, GFN2-xTB, and their r<sup>2</sup>SCAN-3c:GFN2-xTB ONIOM combination with the TPSS-D4/def2-SVP optimized geometries implicitly solvated in toluene

# DALTES/ $\omega$ B97-3c

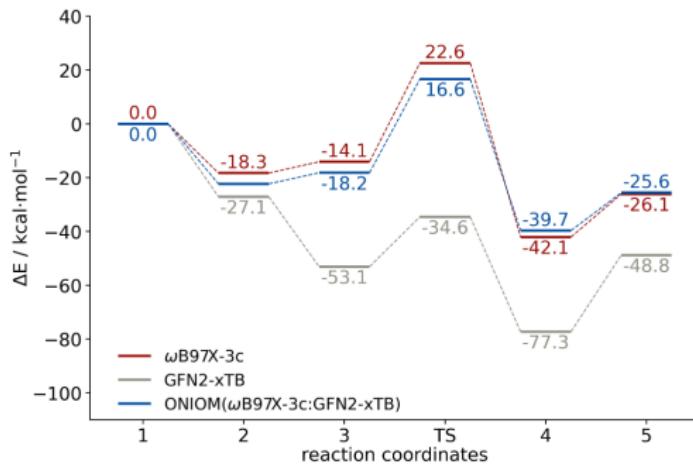


Figure: The relative potential energy curve of the cyanosilylation reaction computed with  $\omega$ B97-3c, GFN2-xTB, and their  $\omega$ B97-3c:GFN2-xTB ONIOM combination with the TPSS-D4/def2-SVP optimized geometries implicitly solvated in toluene.

# DALTES/solvation

Table: The cyanosilylation relative reaction energies at the DALTES dirhodium metal node calculated with  $r^2$ SCAN-3c, TPSS-D4/def2-SVP and their ONIOM combinations with GFN2-xTB with the TPSS-D4/def2-SVP optimized geometries. The individual values are given in  $\text{kcal}\cdot\text{mol}^{-1}$ .

Methods	Solvation model	1	2	3	TS	4	5
TPSS-D4/def2-SVP	COSMO	0.00	-25.75	-30.78	1.52	-51.88	-28.99
3*TPSS-D4/def2-SVP//GFN2-xTB	GBSA	0.00	-28.52	-33.98	-3.49	-51.39	-25.98
	ddCOSMO	0.00	-28.44	-32.84	-2.04	-47.80	-28.15
	COSMO/ddCOSMO	0.00	-25.69	-36.56	-7.05	-43.54	-36.85
$r^2$ SCAN-3c	COSMO	0.00	-18.42	-20.58	13.28	-44.34	-30.13
3* $r^2$ SCAN-3c//GFN2-xTB	GBSA	0.00	-20.28	-24.20	8.73	-43.54	-25.78
	ddCOSMO	0.00	-21.97	-24.89	8.41	-41.72	-29.78
	COSMO/ddCOSMO	0.00	-27.36	-27.58	3.29	-60.00	-37.12



# UiO-66/Ancpt

Methods	size	hRMSD(vs X-ray) / Å	Wall-time (14cores)
GFN-FF	full	0.335	5s
3*GFN2-xTB//GFN-FF	1	0.290	12s
	3	0.233	49s
	6	0.115	1m 54s
3*TPSS-D4/def2-SVP//GFN-FF	1	0.288	52m
	3	0.213	5h 55m
	6	0.025	11h 1m
3*r <sup>2</sup> SCAN-3c//GFN-FF	1	0.292	2h 21m
	3	0.219	12h 50m
	6	0.025	20h 57m
3*ωB97X-3c//GFN-FF	1	0.291	16h 16m
	3	0.212	113h 19m
	6	0.031	202h 51m
GFN2-xTB	full	0.156	3m
3*TPSS-D4/def2-SVP//GFN2-xTB	1	0.155	1h 30m
	3	0.119	6h 1m
	6	0.059	10h 45m
3*r <sup>2</sup> SCAN-3c//GFN2-xTB	1	0.161	3h 31m
	3	0.124	19h 22m
	6	0.054	18h 20m
3*ωB97X-3c//GFN2-xTB	1	0.158	31h 35m
	3	0.126	169h 59m
	6	0.069	334h 32m
TPSS-D4/def2-SVP	full	0.031	46h 9m
r <sup>2</sup> SCAN-3c	full	0.030	190h 38m
ωB97X-3c	full	-	-



# UiO-66/FIRE

Methods	size	hRMSD(vs X-ray) / Å	Wall-time (14cores)
GFN-FF	full	0.285	16s
3*GFN2-xTB//GFN-FF	1	0.273	36s
	3	0.199	1m 56s
	6	0.111	6m 24s
3*TPSS-D4/def2-SVP//GFN-FF	1	0.264	2h 58m
	3	0.191	18h 18m
	6	0.022	30h 38m
3*r <sup>2</sup> SCAN-3c//GFN-FF	1	0.269	5h 50m
	3	0.190	28h 24m
	6	0.023	62h 56m
3*ωB97X-3c//GFN-FF	1	0.267	52h 13m
	3	0.151	213h
	6	0.031	915h 30m
GFN2-xTB	full	0.143	8m
3*TPSS-D4/def2-SVP//GFN2-xTB	1	0.133	2h 24m
	3	0.111	16h 9m
	6	0.050	34h 44m
3*r <sup>2</sup> SCAN-3c//GFN2-xTB	1	0.134	4h 15m
	3	0.109	25h 18m
	6	0.049	71h 58m
3*ωB97X-3c//GFN2-xTB	1	0.140	44h 11m
	3	0.111	269h 44m
	6	0.069	707h 13m
TPSS-D4/def2-SVP	full	0.031	46h 9m
r <sup>2</sup> SCAN-3c	full	0.030	190h 38m
ωB97X-3c	full	-	-

# Purification

$$E = 2 \sum_{i=1}^N \epsilon_i + V_{dc} = Tr[PH] + V_{dc} \quad (8)$$

GCP:

$$\alpha = \min \left\{ \frac{1}{H_{max} - \mu}, \frac{1}{\mu - H_{min}} \right\} \quad (9)$$

Gershgorin's circle theorem:

$$\epsilon_{max} \geq \max_i \left\{ H_{ii} + \sum_{i \neq j} \|H_{ij}\| \right\} \quad (10)$$

$$\epsilon_{min} \leq \min_i \left\{ H_{ii} - \sum_{i \neq j} \|H_{ij}\| \right\} \quad (11)$$

$$\gamma = \frac{Tr[p_k Sp_k - p_k Sp_k Sp_k]}{Tr[p_k - p_k Sp_k]} \quad (12)$$