



# SEMINAR 2:

## THE DEEPMIND 2021 FUNCTIONAL

Simulating matter on the quantum scale with AI

Kyle R. Bryenton<sup>1</sup>

<sup>1</sup>Dalhousie Department of Physics & Atmospheric Science

February 01, 2022  
PHYC6802: Phys&Atmos PhD Seminar Series

# THE DEEPMIND 2021 FUNCTIONAL

## SIMULATING MATTER ON THE QUANTUM SCALE WITH AI

### ABSTRACT

The N-electron problem may be greatly simplified by writing the system's total energy as a functional of electron density. This density-functional theory (DFT) has become the main workhorse of condensed-matter physics and quantum chemistry, enabling first-principles computational modelling of the electronic structure of large-scale quantum systems. Unfortunately, there is no free lunch. While DFT is known to efficiently and accurately simulate many systems of interest, it is also known to catastrophically fail with others. Two of the largest classes of systems that fail are those with fractional spin and fractional charge. These classes of systems are considered to be two of the largest outstanding challenges in DFT.

Introduced in a recently-published Science article, the DeepMind research group has designed a new density functional (DM21) to tackle both aforementioned systems. By training a deep neural network on accurate chemical data and fractional electron constraints, the resulting functional demonstrates the potential to compete with the best (and computationally most expensive) functionals in the literature on both fractional-spin and fractional-charge systems.

This seminar will introduce the basic principles behind Density-Functional Theory and neural networks before exploring the methodology and results of the article: Kirkpatrick, James, et al. "Pushing the frontiers of density functionals by solving the fractional electron problem." *Science* 374.6573 (2021). doi: 10.1126/science.abj6511

Density-Functional Theory  
ooooo

DeepMind & Neural Networks  
ooo

The DeepMind 2021 Functional  
oooooooo

Conclusion  
ooo

**Density Functional Theory**  
**Hohenberg-Kohn-Sham DFT**  
**Fractional Charge & Fractional Spin**

**DeepMind & Neural Networks**  
**What Is DeepMind?**  
**Neural Networks**

**The DeepMind 2021 Functional**  
**Architecture & Training**  
**Applications & Benchmarks**

**Conclusions**

Density-Functional Theory

ooooo

DeepMind & Neural Networks

ooo

The DeepMind 2021 Functional

oooooooo

Conclusion

ooo

## Density Functional Theory

Hohenberg-Kohn-Sham DFT

Fractional Charge & Fractional Spin

## DeepMind & Neural Networks

## The DeepMind 2021 Functional

## Conclusions

## HOHENBERG-KOHN-SHAM DFT

Density-functional theory (DFT) has proven to be one of the most accurate ways of calculating the electronic structure of molecules and solids while maintaining sufficiently low computational requirements.

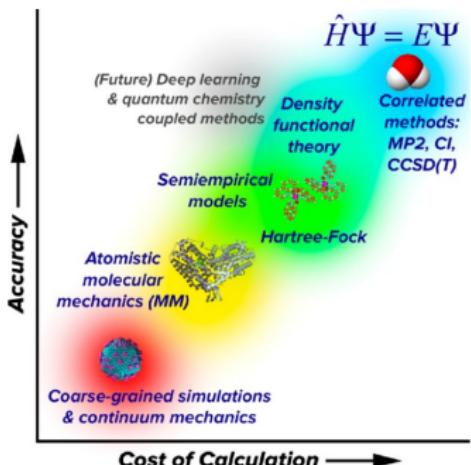
The  $N$ -electron problem is greatly simplified by writing the total energy,

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}\rho(\mathbf{r})d\mathbf{r} + E_H[\rho] + E_{\text{xc}}[\rho],$$

as a functional of the electron density,

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2,$$

for  $N$  electrons occupying orbitals  $\psi_i(\mathbf{r})$ .



## ENERGY FUNCTIONAL COMPONENTS

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + E_H[\rho] + E_{XC}[\rho]$$

### The Kohn-Sham Kinetic Energy

$$T_s[\rho] = \sum_{i=1}^N \left\langle \psi_i(\mathbf{r}) \left| -\frac{\hbar^2}{2m} \nabla^2 \right| \psi_i(\mathbf{r}) \right\rangle$$

Via the Born-Oppenheimer approximation, the nuclei's kinetic energy is zero.

### The External Potential

$v_{\text{ext}}(\mathbf{r})$  is the external potential that is acting on the system.

### The Hartree Energy

$$E_H[\rho] = \frac{e^2}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$$

Also known as the Coulombic energy, sometimes denoted  $J[\rho]$ .

### The Exchange-Correlation Energy

This exchange-correlation functional is chosen by the researcher. In our case:

$$E_{XC}[\rho] = E_{XC}^{\text{DM21}}[\rho].$$

While the exact (or universal) functional is theorized to exist, the explicit form is unknown.

## ENERGY FUNCTIONAL COMPONENTS

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + E_H[\rho] + E_{XC}[\rho]$$

### The Kohn-Sham Kinetic Energy

$$T_s[\rho] = \sum_{i=1}^N \left\langle \psi_i(\mathbf{r}) \left| -\frac{\hbar^2}{2m} \nabla^2 \right| \psi_i(\mathbf{r}) \right\rangle$$

Via the Born-Oppenheimer approximation, the nuclei's kinetic energy is zero.

### The External Potential

$v_{\text{ext}}(\mathbf{r})$  is the external potential that is acting on the system.

### The Hartree Energy

$$E_H[\rho] = \frac{e^2}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$$

Also known as the Coulombic energy, sometimes denoted  $J[\rho]$ .

### The Exchange-Correlation Energy

This exchange-correlation functional is chosen by the researcher. In our case:

$$E_{XC}[\rho] = E_{XC}^{\text{DM21}}[\rho].$$

While the exact (or universal) functional is theorized to exist, the explicit form is unknown.

## ENERGY FUNCTIONAL COMPONENTS

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + E_H[\rho] + E_{XC}[\rho]$$

### The Kohn-Sham Kinetic Energy

$$T_s[\rho] = \sum_{i=1}^N \left\langle \psi_i(\mathbf{r}) \left| -\frac{\hbar^2}{2m} \nabla^2 \right| \psi_i(\mathbf{r}) \right\rangle$$

Via the Born-Oppenheimer approximation, the nuclei's kinetic energy is zero.

### The External Potential

$v_{\text{ext}}(\mathbf{r})$  is the external potential that is acting on the system.

### The Hartree Energy

$$E_H[\rho] = \frac{e^2}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$$

Also known as the Coulombic energy, sometimes denoted  $J[\rho]$ .

### The Exchange-Correlation Energy

This exchange-correlation functional is chosen by the researcher. In our case:

$$E_{XC}[\rho] = E_{XC}^{\text{DM21}}[\rho].$$

While the exact (or universal) functional is theorized to exist, the explicit form is unknown.

## ENERGY FUNCTIONAL COMPONENTS

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + E_H[\rho] + E_{XC}[\rho]$$

### The Kohn-Sham Kinetic Energy

$$T_s[\rho] = \sum_{i=1}^N \left\langle \psi_i(\mathbf{r}) \left| -\frac{\hbar^2}{2m} \nabla^2 \right| \psi_i(\mathbf{r}) \right\rangle$$

Via the Born-Oppenheimer approximation, the nuclei's kinetic energy is zero.

### The External Potential

$v_{\text{ext}}(\mathbf{r})$  is the external potential that is acting on the system.

### The Hartree Energy

$$E_H[\rho] = \frac{e^2}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$$

Also known as the Coulombic energy, sometimes denoted  $J[\rho]$ .

### The Exchange-Correlation Energy

This exchange-correlation functional is chosen by the researcher. In our case:

$$E_{XC}[\rho] = E_{XC}^{\text{DM21}}[\rho].$$

While the exact (or universal) functional is theorized to exist, the explicit form is unknown.

## ENERGY FUNCTIONAL COMPONENTS

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + E_H[\rho] + E_{XC}[\rho]$$

### The Kohn-Sham Kinetic Energy

$$T_s[\rho] = \sum_{i=1}^N \left\langle \psi_i(\mathbf{r}) \left| -\frac{\hbar^2}{2m} \nabla^2 \right| \psi_i(\mathbf{r}) \right\rangle$$

Via the Born-Oppenheimer approximation, the nuclei's kinetic energy is zero.

### The External Potential

$v_{\text{ext}}(\mathbf{r})$  is the external potential that is acting on the system.

### The Hartree Energy

$$E_H[\rho] = \frac{e^2}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$$

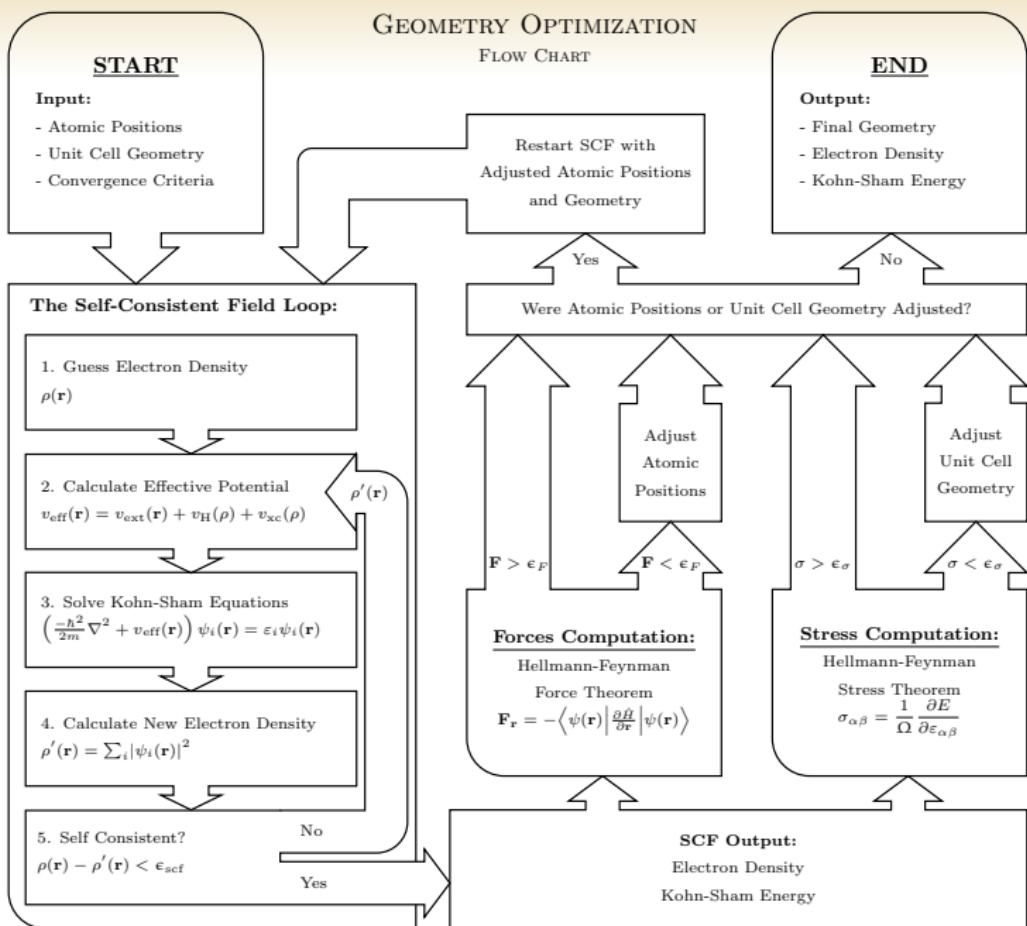
Also known as the Coulombic energy, sometimes denoted  $J[\rho]$ .

### The Exchange-Correlation Energy

This exchange-correlation functional is chosen by the researcher. In our case:

$$E_{XC}[\rho] = E_{XC}^{\text{DM21}}[\rho].$$

While the exact (or universal) functional is theorized to exist, the explicit form is unknown.



## FRACTIONAL CHARGE

- Error that is caused by highly delocalized electron densities
- Better known as “Delocalization Error” or “Self-Interaction Error”
- Considered the greatest outstanding challenge of DFT
- Most notable example: stretched  $\text{H}_2^+$

## FRACTIONAL SPIN

- Unrestricted calculations allow partial orbital occupation
- Linked to in multi-reference/strong-correlation error
- Not specific to DFT, present in wavefunction methods
- Most notable example: stretched  $\text{H}_2$

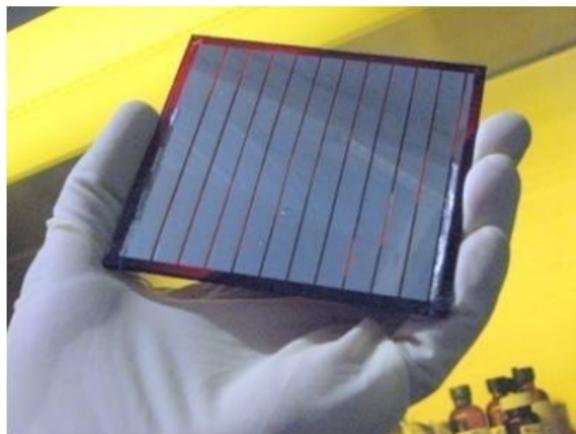
# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems



---

[https://en.wikipedia.org/wiki/Organic\\_solar\\_cell](https://en.wikipedia.org/wiki/Organic_solar_cell)

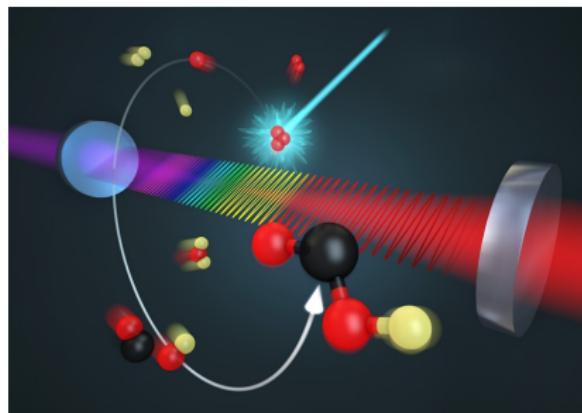
# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems



(The Ye group and Steve Burrows / JILA)

<https://jila.colorado.edu/news-events/articles/radical-comb-over>

# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems



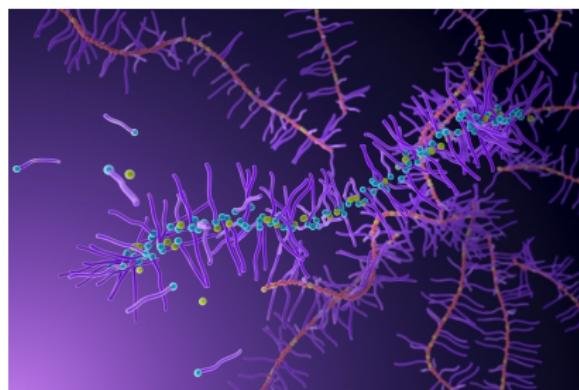
---

[https://www.adlinktech.com/en/  
semiconductor-solution-packaging-and-testing](https://www.adlinktech.com/en/semiconductor-solution-packaging-and-testing)

# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds



## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems

(Demin Liu)

[https://news.mit.edu/2019/  
new-synthesis-method-yields-degradable-polymers-1028](https://news.mit.edu/2019/new-synthesis-method-yields-degradable-polymers-1028)

# EXAMPLES

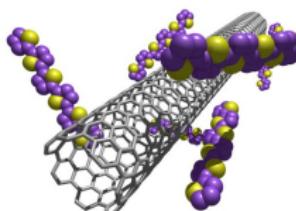
## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

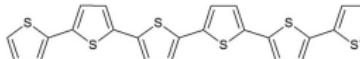
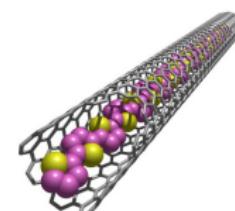
## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems

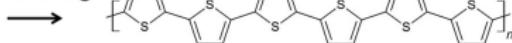
### Vapor-phase encapsulation



### Polymerization



### Annealing



Miyaura *et al.* Sci Rep 8, 8098 (2018) doi:  
10.1038/s41598-018-26379-4

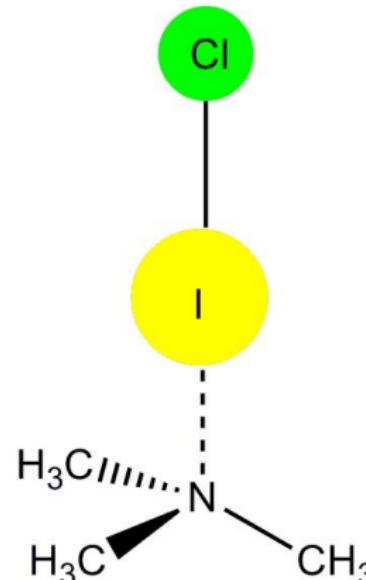
# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems



[https://en.wikipedia.org/wiki/Halogen\\_bond](https://en.wikipedia.org/wiki/Halogen_bond)

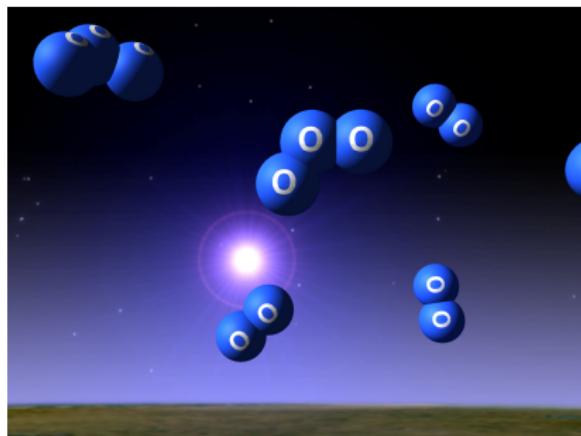
# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems

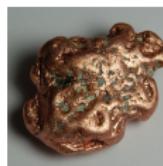


NASA/Goddard Space Flight Center Scientific  
Visualization Studio  
<https://svs.gsfc.nasa.gov/823>

# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds



(a)



(b)



(c)

---

Theopold *et al.* Properties of Transition Metals and Their Compounds,  
<https://chem.libretexts.org/@go/page/38321>

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems

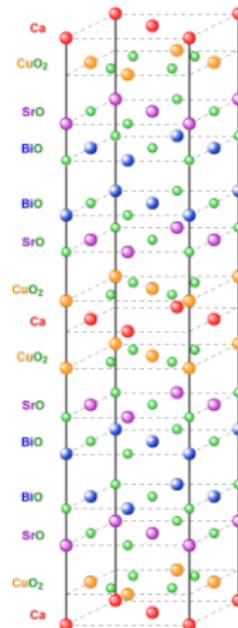
# EXAMPLES

## Fractional Charge:

- Charge-transfer complexes
- Transition states of radical reactions
- Band gaps of semi-conductors
- Polarizabilities of long-chain molecules
- Systems with extended conjugation
- Halogen/Chalcogen bonds

## Fractional Spin:

- Biradicals
- Some transition-metal compounds
- Strongly-correlated systems



(BSCCO) [https://en.wikipedia.org/wiki/Strongly\\_correlated\\_material](https://en.wikipedia.org/wiki/Strongly_correlated_material)

Density-Functional Theory  
ooooo

DeepMind & Neural Networks  
ooo

The DeepMind 2021 Functional  
oooooooo

Conclusion  
ooo

## Density Functional Theory

## DeepMind & Neural Networks

What Is DeepMind?

Neural Networks

## The DeepMind 2021 Functional

## Conclusions

## WHAT ARE THEY KNOWN FOR?

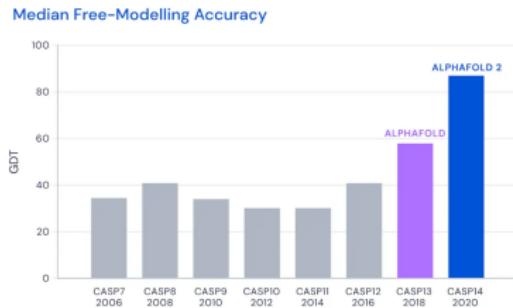
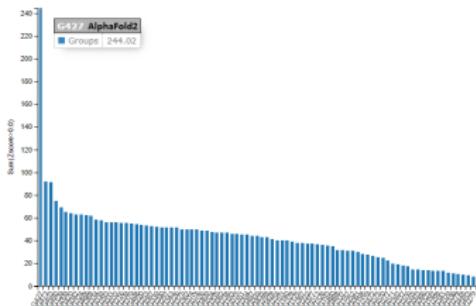
- 2014 Google acquires DeepMind for US\$650 million
- 2015 AlphaGo beats Go champion Fan Hui
- 2016 WaveNet is created, powering Google assistant
- 2016 Google servers cut energy consumption by 15%
- 2018 Identifying 50+ eye diseases with 95% accuracy
- 2020 AlphaFold2 solves protein folding

T1037 / 6vr4  
90.7 GDT  
(RNA polymerase domain)

T1049 / 6y4f  
93.3 GDT  
(adhesin tip)

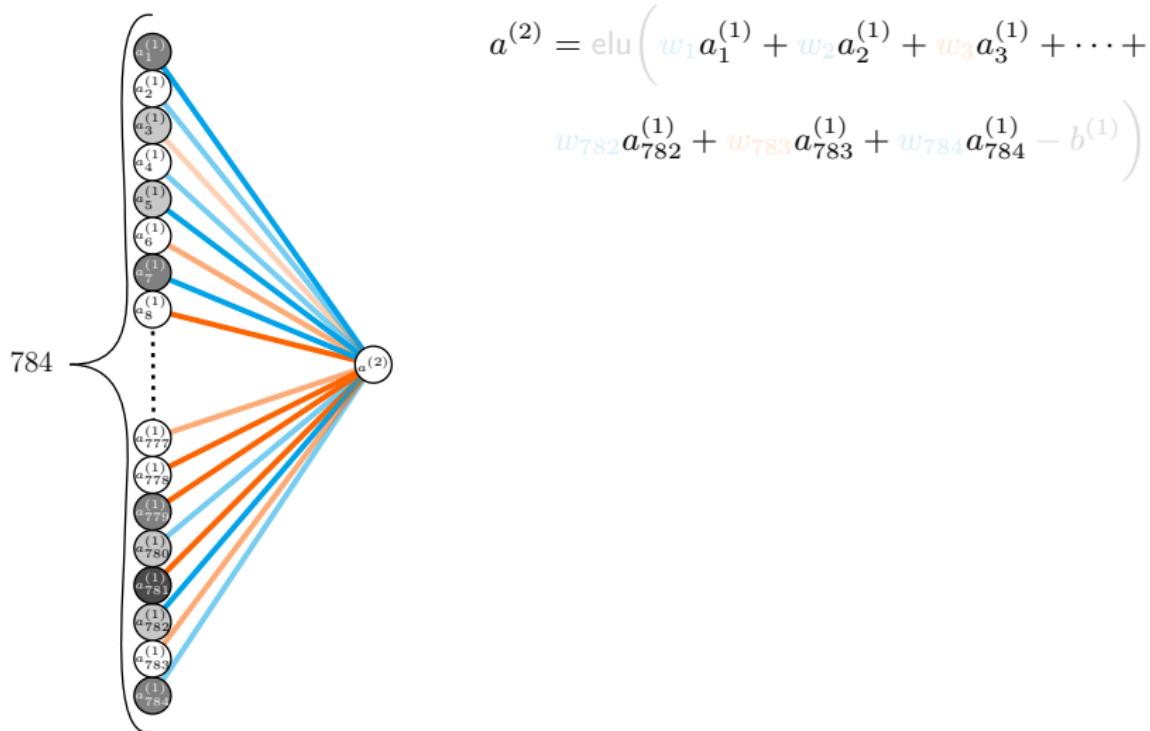
● Experimental result  
● Computational prediction

## Critical Assessment of protein Structure Prediction (CASP)

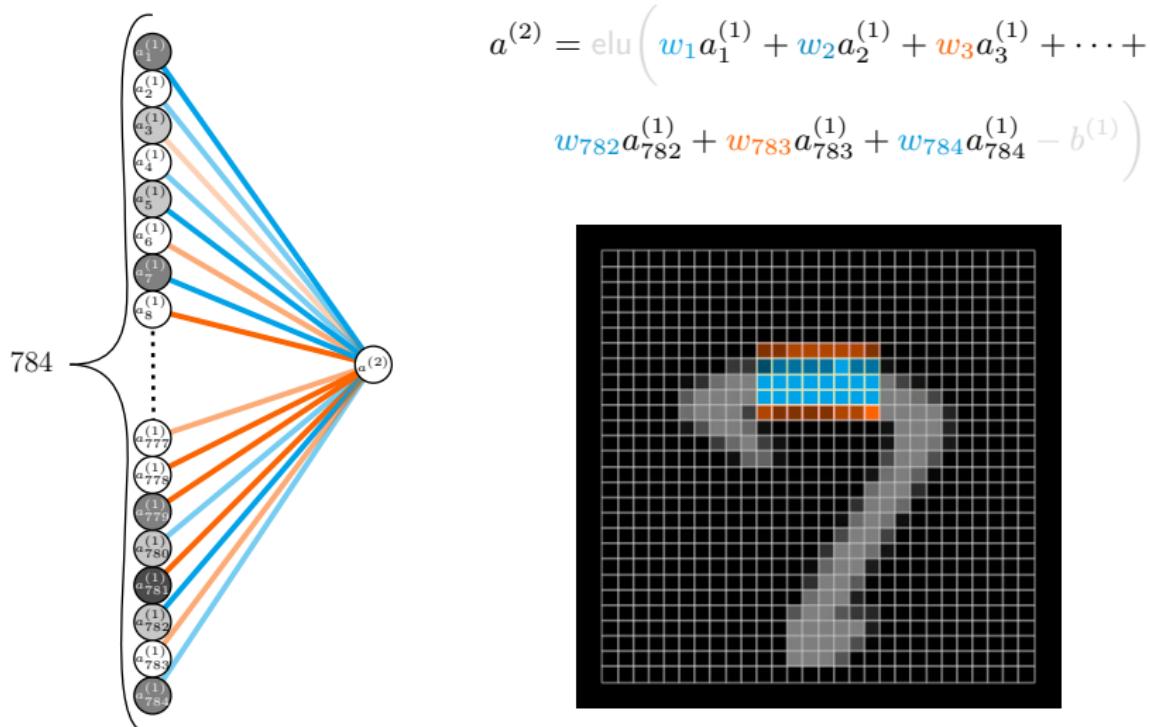


# WHAT IS A NEURAL NETWORK?

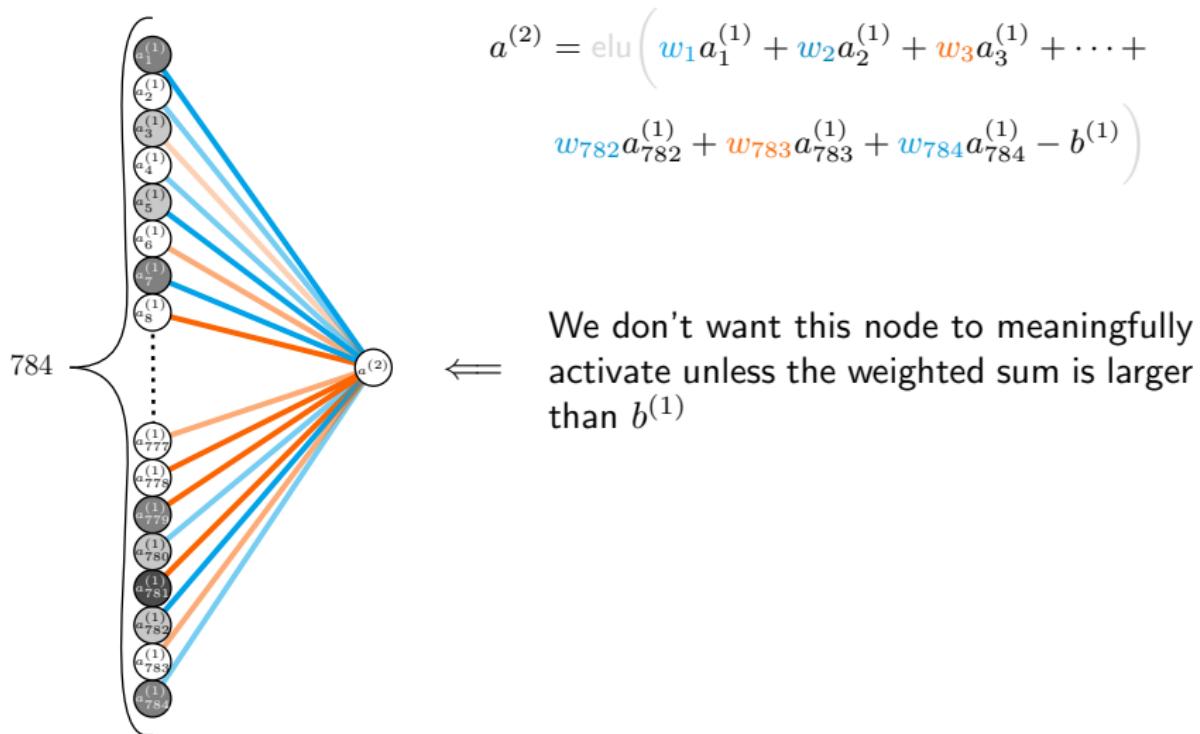
## EXAMPLE NODE COMPUTATION



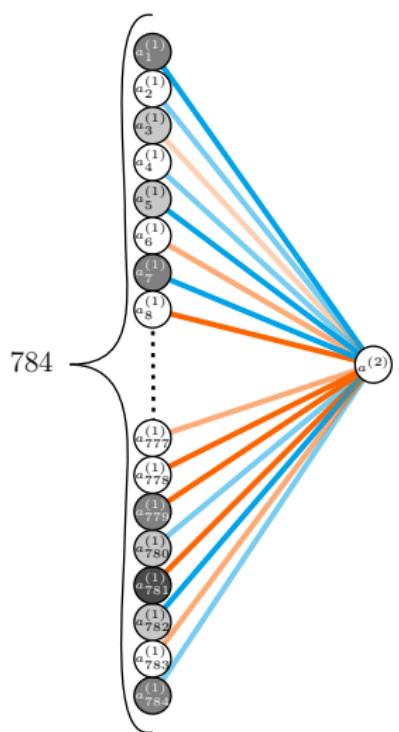
## EXAMPLE NODE COMPUTATION



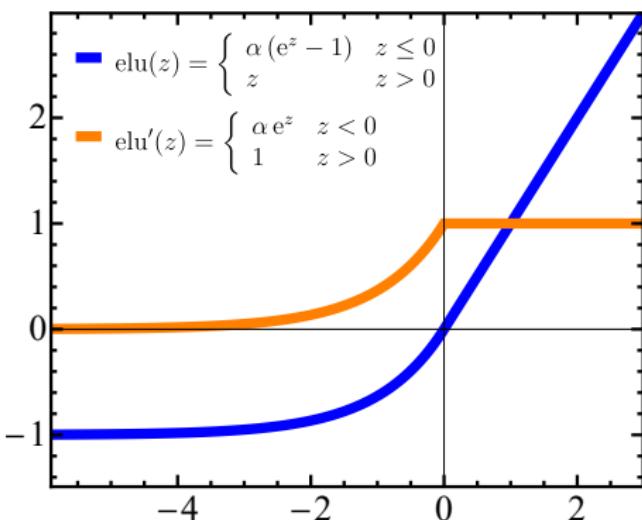
## EXAMPLE NODE COMPUTATION



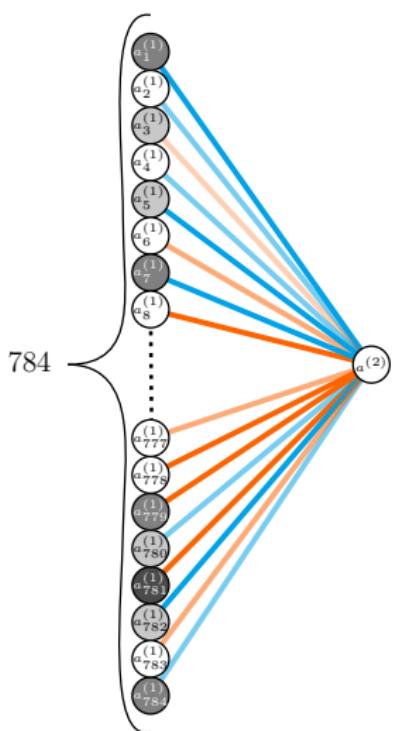
## EXAMPLE NODE COMPUTATION



$$a^{(2)} = \text{elu} \left( w_1 a_1^{(1)} + w_2 a_2^{(1)} + w_3 a_3^{(1)} + \dots + w_{782} a_{782}^{(1)} + w_{783} a_{783}^{(1)} + w_{784} a_{784}^{(1)} - b^{(1)} \right)$$



## EXAMPLE NODE COMPUTATION



$$a^{(2)} = \text{elu} \left( w_1 a_1^{(1)} + w_2 a_2^{(1)} + w_3 a_3^{(1)} + \dots + w_{782} a_{782}^{(1)} + w_{783} a_{783}^{(1)} + w_{784} a_{784}^{(1)} - b^{(1)} \right)$$

The neural network is “trained” via *backpropagation*, where the biases, weights, and previous layer’s nodes are shifted in the direction of steepest descent with respect to the *loss function*.

$$\mathcal{L} = \sum_j \left( a_j^{(L)} - y_j \right)^2$$

$$\frac{\partial \mathcal{L}}{\partial b_j^{(L)}} \quad \frac{\partial \mathcal{L}}{\partial w_{jk}^{(L)}} \quad \frac{\partial \mathcal{L}}{\partial a_k^{(L-1)}}$$

Density-Functional Theory  
ooooo

DeepMind & Neural Networks  
ooo

The DeepMind 2021 Functional  
oooooooo

Conclusion  
ooo

Density Functional Theory

DeepMind & Neural Networks

The DeepMind 2021 Functional  
Architecture & Training  
Applications & Benchmarks

Conclusions

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $x(r)$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained

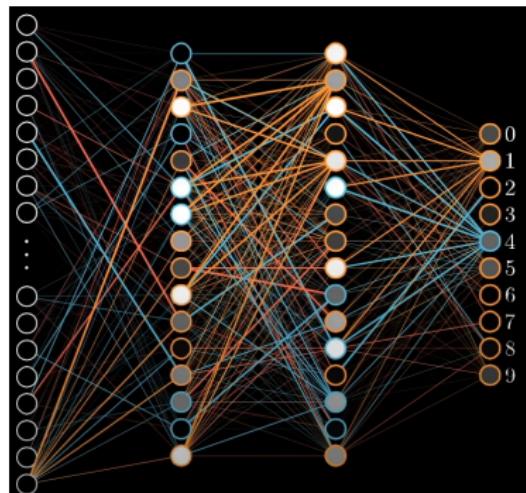


---

[https://characterprofile.fandom.com/wiki/Megatron\\_\(G1\)](https://characterprofile.fandom.com/wiki/Megatron_(G1))

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $x(r)$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained



3Blue1Brown - What is backpropagation really doing? — Chapter 3, Deep learning

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $x(r)$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained

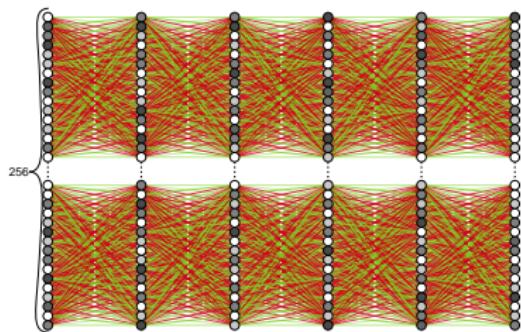
$$E_{XC}^{DM21} = E_{XC}^{\text{MLP}} + E_{D3(\text{BJ})}$$

$$E_{D3(\text{BJ})} = - \sum_{\text{Pairs}} \sum_{n=6,8} s_n \frac{C_n}{R^n} f_n^{\text{BJ}}(R)$$

$$f_n^{\text{BJ}}(R) = \frac{R^n}{R^n + R_{\text{vdW}}^n}$$

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $x(r)$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained



## FUNCTIONAL &amp; NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $x(r)$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained

$$\begin{aligned} \mathbf{w}^{(L)} &= \begin{bmatrix} w_{1,1}^{(L)} & w_{1,2}^{(L)} & \cdots & w_{1,n_L}^{(L)} \\ w_{2,1}^{(L)} & w_{2,2}^{(L)} & \cdots & w_{2,n_L}^{(L)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_L,1}^{(L)} & w_{n_L,2}^{(L)} & \cdots & w_{n_L,n_L}^{(L)} \end{bmatrix} \\ &\approx \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \end{aligned}$$

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $\mathbf{x}(\mathbf{r})$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained

$$\mathbf{x}(\mathbf{r}) = \begin{bmatrix} \rho_{\uparrow} \\ \rho_{\downarrow} \\ |\nabla\rho_{\uparrow}| \\ |\nabla\rho_{\downarrow}| \\ |\nabla\rho| \\ \tau_{\uparrow} \\ \tau_{\downarrow} \\ e_{HF}^{\uparrow} \\ e_{HF}^{\downarrow} \\ e_{\omega HF}^{\uparrow} \\ e_{\omega HF}^{\downarrow} \end{bmatrix}$$

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer ( $\sim 400k$  parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $\mathbf{x}(\mathbf{r})$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained

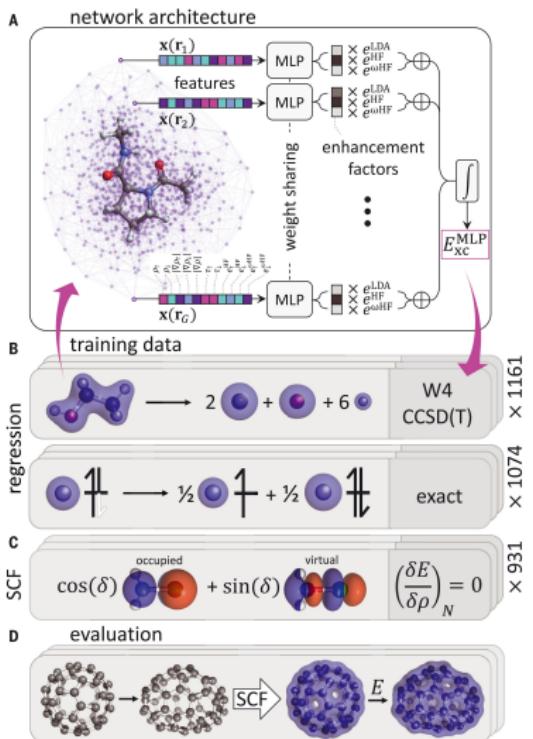
$$\mathbf{x}(\mathbf{r}) = \begin{bmatrix} \rho_{\downarrow} \\ \rho_{\uparrow} \\ |\nabla\rho_{\downarrow}| \\ |\nabla\rho_{\uparrow}| \\ |\nabla\rho| \\ \tau_{\downarrow} \\ \tau_{\uparrow} \\ e_{\downarrow}^{\text{HF}} \\ e_{\uparrow}^{\text{HF}} \\ e_{\downarrow}^{\omega\text{HF}} \\ e_{\uparrow}^{\omega\text{HF}} \end{bmatrix}$$

# FUNCTIONAL & NETWORK ARCHITECTURE OF DM21

- DM21 uses a Multilayer Perceptron
- Combines a MLP with the D3(BJ) dispersion correction
- The MLP is a 6-layer design with 256 nodes per layer (~400k parameters)
- Weight matrices are initialized close to the identity matrix
- 11-D feature vector,  $\mathbf{x}(\mathbf{r})$ , contains selected electronic structure information
- The network is spin symmetric by running the network twice, one for each spin ordering of inputs
- In total, 4 versions of DM21 were trained

|        | FC | FS | UEG |
|--------|----|----|-----|
| DM21   | ✓  | ✓  |     |
| DM21m  |    |    |     |
| DM21mc |    | ✓  |     |
| DM21mu |    |    | ✓   |

## TRAINING DATA

**A)**

- $\mathbf{x}(\mathbf{r})$  for  $G$  gridpoints goes into a shared MLP.
- The MLP predicts  $e_{XC}^{\text{MLP}}$ , integrated to get  $E_{XC}^{\text{MLP}}$ .

**B)**

- Trained on 1161 basic reactions for small main-group H-Kr molecules
- Extra training on 1074 reactions representing FC and FS densities of H-Ar
- Labels obtained from literature, or CCSD(T)

**C)**

- Perturbation theory gives  $\delta E / \delta \rho$  after one SCF iteration.
- This is trained on an addition 931 systems.
- Gradient training is important so DM21 can be used in normal DFT procedures.

**D)**

- Once trained, the functional can be deployed in self-consistent calculations.

## TRAINING OBJECTIVE

The loss/cost function was the sum of squares of a regression and gradient regularization term.

$$\mathcal{L} = \mathbb{E}_r \left[ \left( \Delta E_{XC,r}^{\text{DM21}} - \Delta E_{XC,r}^* \right)^2 \right] + \lambda \mathbb{E}_s [\delta E_{\text{SCF},s}^2]$$

$\mathbb{E}$  the expectation over the dataset of reactions ( $r$ ) or systems ( $s$ )

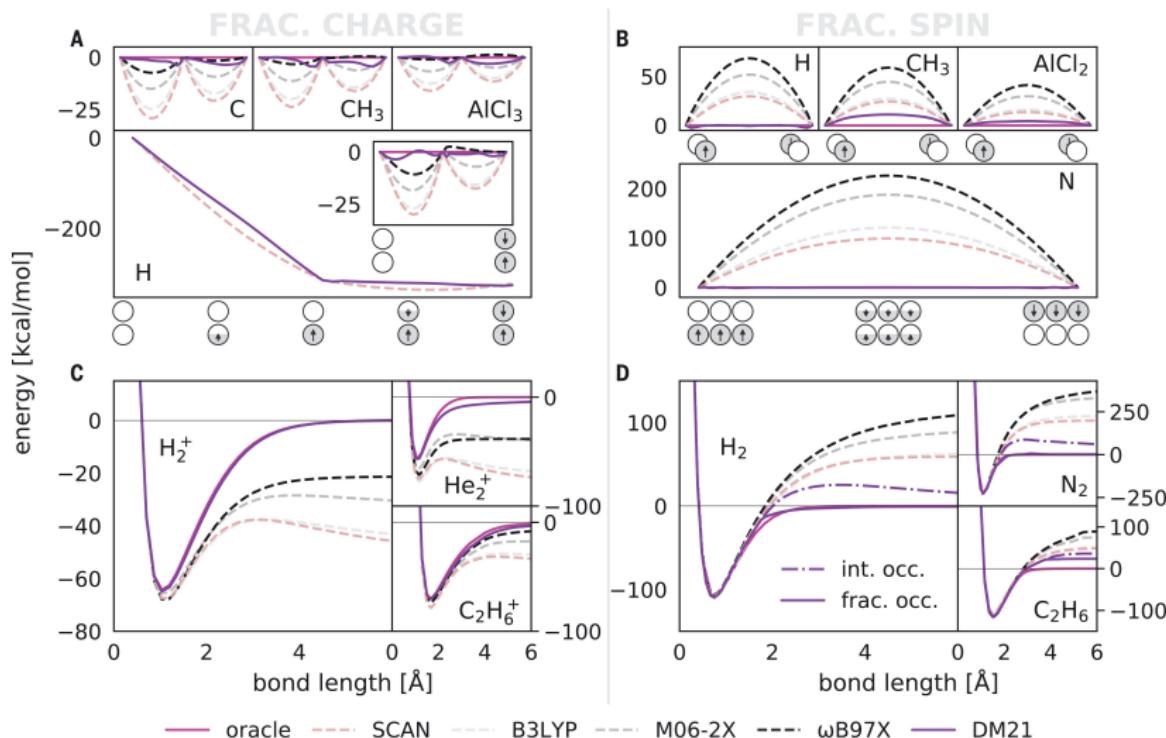
$\lambda$  controls the weight of gradient term (fixed at  $\lambda = 1$ )

$\Delta E_{XC,r}^{\text{DM21}}$  is the model's prediction for the total reaction exchange-correlation energy

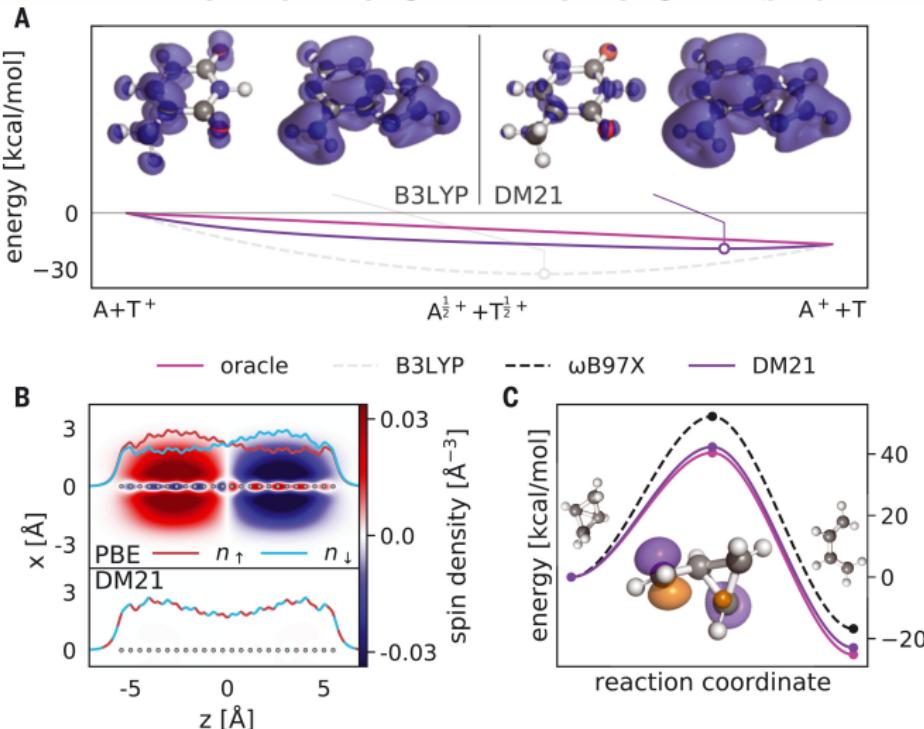
$\Delta E_{XC,r}^*$  is the exchange-correlation energy for the reaction computed from literature or CCSD(T)

$\delta E_{\text{SCF},s}$  gives the leading order change in energy after a single SCF iteration.  
Squaring encourages the model to avoid spuriously large orbital rotations away from reasonable orbitals during the SCF.

## APPLICATION TO TEXTBOOK EXAMPLES

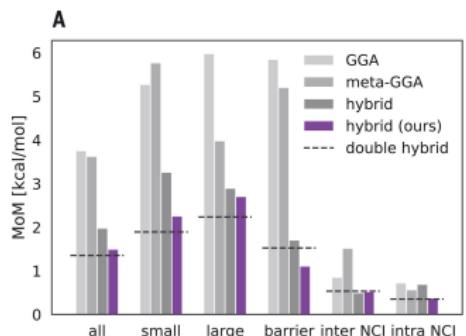


## APPLICATION TO CHALLENGING CHEMISTRY



- A) Charge density for ionized adenine-thymine base pair. (Issue: Charge delocalized on thymine)
- B) Spin density for a compressed chain of 24 hydrogen atoms. (Issue: Broken spin symmetry)
- C) The conrotatory pathways of bicyclobutane isomerization. (Issue: Delocalized spin on path)

## BENCHMARKS



**B**

|                          | GMTKN55     | BBB          | QM9         |
|--------------------------|-------------|--------------|-------------|
|                          | MoM         |              |             |
| SCAN(:D3 <sub>Bj</sub> ) | 3.63        | 52.90        | 3.57        |
| $\omega$ B97X(-V)        | 2.45        | 70.28        | 2.13        |
| M06-2X                   | 2.08        | 64.50        | 2.20        |
| PW6B95:D3 <sub>0</sub>   | 1.98        | 57.81        | 2.24        |
| DM21                     | <b>1.50</b> | <b>13.20</b> | <b>1.66</b> |

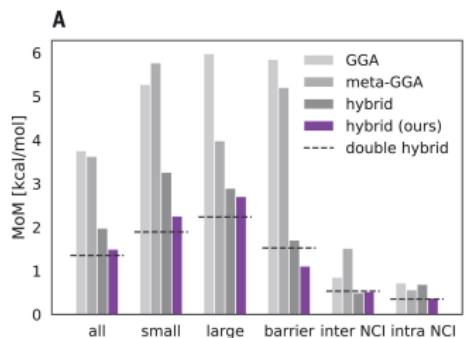
All errors are in kcal/mol, using the mean of means (MoM) error metric.

**GMTKN55** 1505 reactions across 55 sub-benchmarks of general main group thermochemistry, kinetics and noncovalent interactions, includes heavy atoms beyond Kr.

**BBB** A new Quantum Monte Carlo bond breaking benchmark: H<sub>2</sub>, Li<sub>2</sub>, C<sub>2</sub>, N<sub>2</sub>, F<sub>2</sub>

**QM9** A collection of 133,857 enumerated isomers of organic molecules with up to nine heavy atoms.

## BENCHMARKS



**B**

|                          | GMTKN55     | BBB          | QM9         |
|--------------------------|-------------|--------------|-------------|
| MoM                      |             |              |             |
| SCAN(:D3 <sub>BJ</sub> ) | 3.63        | 52.90        | 3.57        |
| $\omega$ B97X(-V)        | 2.45        | 70.28        | 2.13        |
| M06-2X                   | 2.08        | 64.50        | 2.20        |
| PW6B95:D3 <sub>0</sub>   | 1.98        | 57.81        | 2.24        |
| DM21                     | <b>1.50</b> | <b>13.20</b> | <b>1.66</b> |

All errors are in kcal/mol, using the mean of means (MoM) error metric.

**A)** The MoM error metric in each class of reactions from GMTKN55. DM21 is compared with functionals at rungs two to five, with strong GMTKN55 metrics from

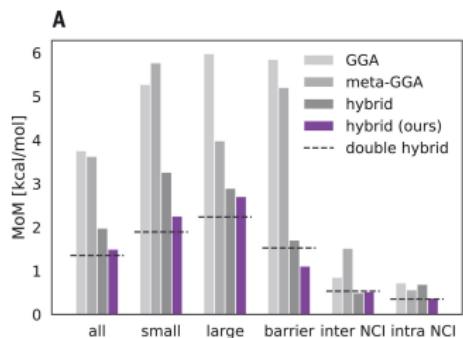
GGA revPBE:D3<sub>BJ</sub>

meta-GGA SCAN:D3<sub>BJ</sub>,

Hybrid PW6B95:D3<sub>0</sub>

Double Hybrid DSD-PBEP86:D3<sub>BJ</sub>

## BENCHMARKS



**B**

|                          | GMTKN55     | BBB          | QM9         |
|--------------------------|-------------|--------------|-------------|
|                          | MoM         |              |             |
| SCAN(:D3 <sub>Bj</sub> ) | 3.63        | 52.90        | 3.57        |
| $\omega$ B97X(-V)        | 2.45        | 70.28        | 2.13        |
| M06-2X                   | 2.08        | 64.50        | 2.20        |
| PW6B95:D3 <sub>0</sub>   | 1.98        | 57.81        | 2.24        |
| DM21                     | <b>1.50</b> | <b>13.20</b> | <b>1.66</b> |

All errors are in kcal/mol, using the mean of means (MoM) error metric.

**B)** Performance of DM21 compared with the SCAN functional and the three best performing hybrid functionals on three benchmark sets. Errors are taken relative to UCCSD(T), QMC, or G4(MP2) levels of theory.

## QUESTIONS I MAY AS WELL ANSWER AHEAD OF TIME:

### HOW DO WE KNOW DM21 DIDN'T MEMORIZE THE RESULTS?

- 1 The functional was trained only on the exact conditions for bare atoms, but correct behavior was also seen on fragments of molecules for both FC and FS.
- 2 DM21 excels at describing barrier heights, but was given no transition states in the training data.
- 3 DM21 outperforms existing functionals on the MB16-43 benchmark subset of GMTKN55, which is designed to test performance with exotic geometries. This dataset was as far from the training and validation set as possible.

## QUESTIONS I MAY AS WELL ANSWER AHEAD OF TIME:

### HOW DO WE KNOW DM21 DIDN'T MEMORIZE THE RESULTS?

- 1 The functional was trained only on the exact conditions for bare atoms, but correct behavior was also seen on fragments of molecules for both FC and FS.
- 2 DM21 excels at describing barrier heights, but was given no transition states in the training data.
- 3 DM21 outperforms existing functionals on the MB16-43 benchmark subset of GMTKN55, which is designed to test performance with exotic geometries. This dataset was as far from the training and validation set as possible.

## QUESTIONS I MAY AS WELL ANSWER AHEAD OF TIME:

### HOW DO WE KNOW DM21 DIDN'T MEMORIZE THE RESULTS?

- 1 The functional was trained only on the exact conditions for bare atoms, but correct behavior was also seen on fragments of molecules for both FC and FS.
- 2 DM21 excels at describing barrier heights, but was given no transition states in the training data.
- 3 DM21 outperforms existing functionals on the MB16-43 benchmark subset of GMTKN55, which is designed to test performance with exotic geometries. This dataset was as far from the training and validation set as possible.

## QUESTIONS I MAY AS WELL ANSWER AHEAD OF TIME:

## WHY DIDN'T THEY ALSO ML THE DISPERSION CORRECTION?

They tried to, but couldn't capture the required long-range behaviour with their set of features.

|                 | GMTKN55( $\leq Kr$ ) |         |         |
|-----------------|----------------------|---------|---------|
|                 | MoM                  | WTMAD-1 | WTMAD-2 |
| $\omega$ B97X-V | 2.29                 | 2.34    | 3.89    |
| DM21            | 1.50                 | 1.96    | 3.81    |
| DM21m           | 1.27                 | 2.11    | 3.76    |
| DM21mc          | 1.36                 | 2.09    | 3.70    |
| DM21:no-D3      | 1.42                 | 2.70    | 5.33    |
| DM21m:no-D3     | 1.48                 | 2.83    | 5.48    |
| DM21mc:no-D3    | 1.50                 | 2.88    | 4.89    |

$$\mathbf{x}(\mathbf{r}) = \begin{bmatrix} \rho_{\uparrow} \\ \rho_{\downarrow} \\ |\nabla \rho_{\uparrow}| \\ |\nabla \rho_{\downarrow}| \\ |\nabla \rho| \\ \tau_{\uparrow} \\ \tau_{\downarrow} \\ e_{HF}^{\uparrow} \\ e_{HF}^{\downarrow} \\ e_{\omega HF}^{\uparrow} \\ e_{\omega HF}^{\downarrow} \end{bmatrix}$$

Density-Functional Theory  
ooooo

DeepMind & Neural Networks  
ooo

The DeepMind 2021 Functional  
oooooooo

Conclusion  
ooo

Density Functional Theory

DeepMind & Neural Networks

The DeepMind 2021 Functional

Conclusions

# SUMMARY

## Praise:

- Great implementation
- Incredible accuracy
- Excellent supplementary data
- Data availability
- Code availability
- The paper told a story

## Critiques:

- Right for the wrong reasons?
- Not all is good (probably).

"For the chemistry of main-group elements, DM21 is very good, although it may be less accurate for transition metal chemistry, a more challenging problem to which it was not fitted. Solids and liquids could also be described unsatisfactorily for several reasons: because they are not included in the DM21 fitting sets... and because atoms and small molecules can be well described by using full exact exchange at long range, whereas extended systems cannot." ~ Prof. John Perdew

# SUMMARY

## Praise:

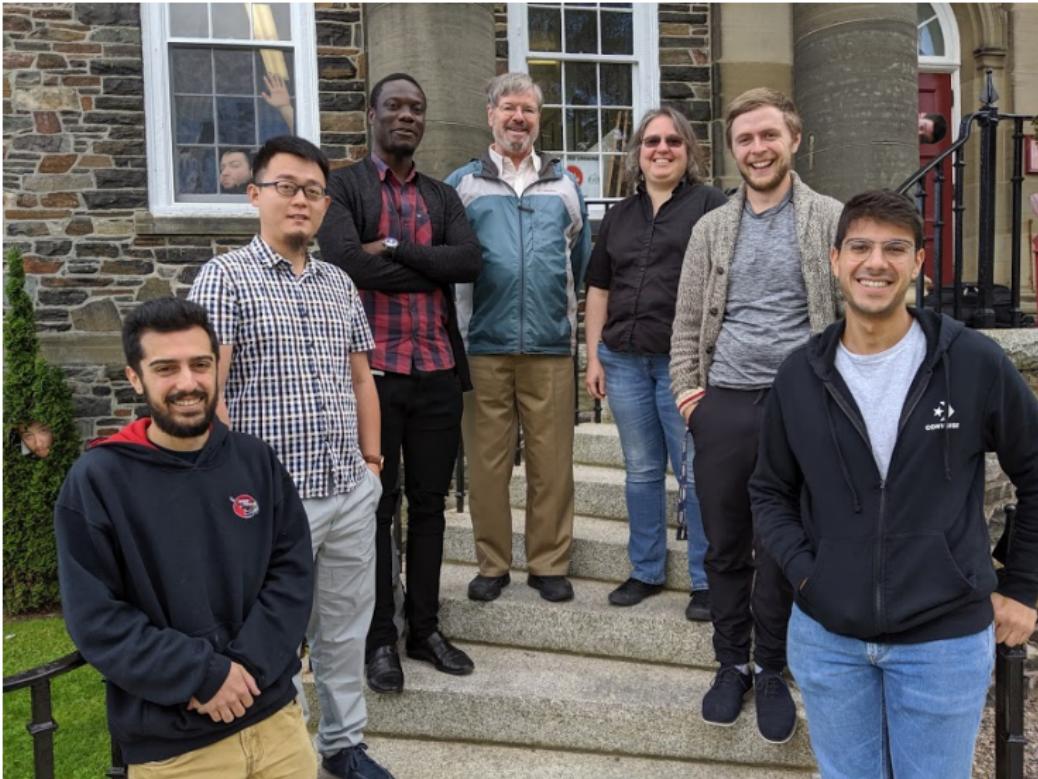
- Great implementation
- Incredible accuracy
- Excellent supplementary data
- Data availability
- Code availability
- The paper told a story

## Critiques:

- Right for the wrong reasons?
- Not all is good (probably).

"For the chemistry of main-group elements, DM21 is very good, although it may be less accurate for transition metal chemistry, a more challenging problem to which it was not fitted. Solids and liquids could also be described unsatisfactorily for several reasons: because they are not included in the DM21 fitting sets... and because atoms and small molecules can be well described by using full exact exchange at long range, whereas extended systems cannot." ~ Prof. John Perdew

# ACKNOWLEDGEMENTS



# QUESTIONS?

WANT MY SLIDES?



KYLE.BRYENTON@DAL.CA

## APPENDIX

### DFT Supplement

The Kohn-Sham Equation

Jacob's Ladder of Density Functional Approximations

The Exchange-Correlation Hole Picture of Delocalization Error

### Neural Network Supplement

That Name Sounds Familiar

Machine Learned Via Backpropagation

### DeepMind Supplement

Training Data (Full)

Benchmarks (Full)

Basis & Code

## THE KOHN-SHAM EQUATION

Once the components of the density energy functional are known, we may derive the effective Kohn-Sham potential for the system

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}[\rho] + v_{\text{xc}}[\rho],$$

where

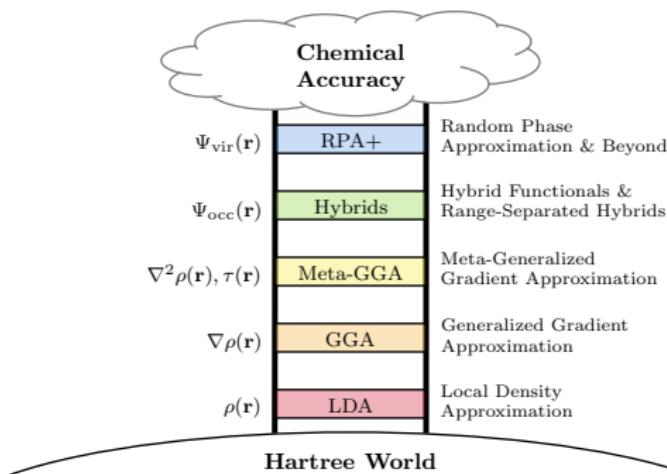
$$v_{\text{H}}[\rho] = e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad v_{\text{xc}}[\rho] = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}.$$

Then the Kohn-Sham equation,

$$\left( -\frac{\hbar^2}{2m} \nabla^2 + v_{\text{eff}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}),$$

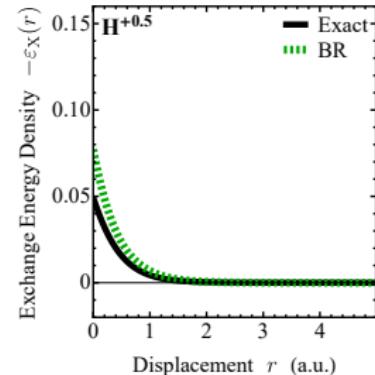
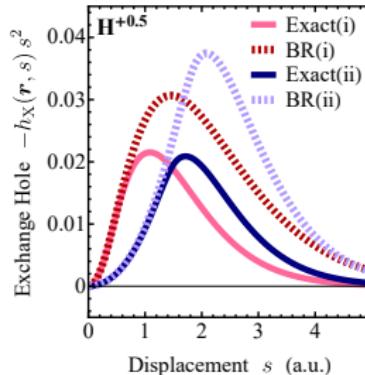
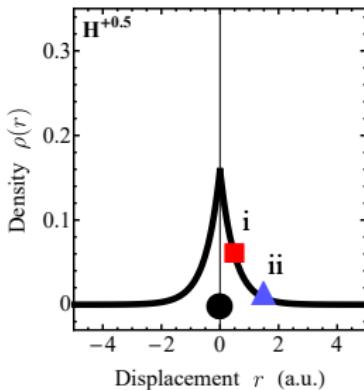
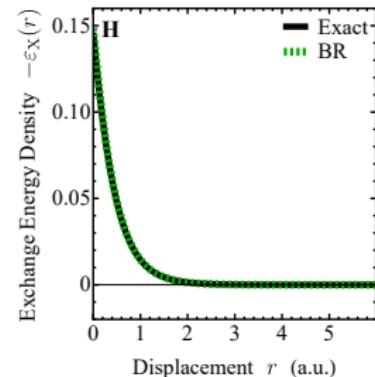
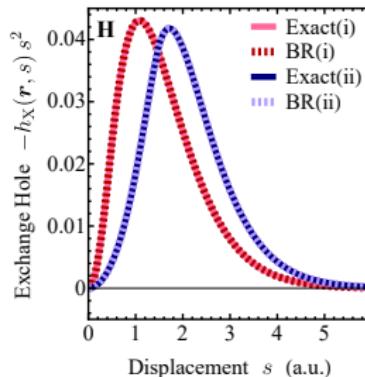
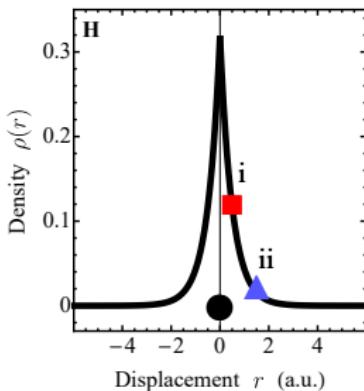
may be solved for its eigenenergies,  $\varepsilon_i$ , and corresponding Kohn-Sham orbital eigenvectors,  $\psi_i(\mathbf{r})$ . This lets us self-consistently calculate the electron density  $\rho(r)$  and the Kohn-Sham energy.

# JACOB'S LADDER OF DENSITY FUNCTIONAL APPROXIMATIONS



- Density-functional theory is functionally exact in its formulation.
- Errors come from DFAs made to  $E_{\text{XC}}[\rho]$ , not the theory itself.
- Ascending leads to better chemical accuracy at increased computational cost.

# THE EXCHANGE-CORRELATION HOLE PICTURE



# THAT NAME SOUNDS FAMILIAR

## Alphabet



# MACHINE LEARNED VIA BACKPROPAGATION

$C_0 \rightarrow$  Cost function

$a_j^{(L)} \rightarrow$  The  $j$ th node's value on layer  $L$

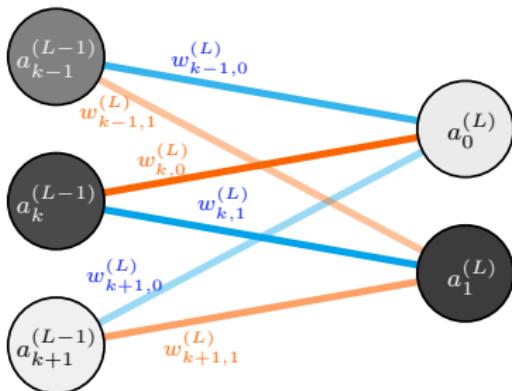
$L \rightarrow$  Layer index

$b_j^{(L)} \rightarrow$  The  $j$ th node's bias on layer  $L$

$y_j \rightarrow$  Desired value of  $a_j^{(L)}$

$w_{jk}^{(L)} \rightarrow$  Weight connecting  $a_k^{(L-1)}$  and  $a_j^{(L)}$

We quantify how close we were to the correct answer via a *cost function*.



$$C_0 = \sum_j (a_j^{(L)} - y_j)^2$$

$$a_j^{(L)} = \sigma(z_j^{(L)})$$

$$z_j^{(L)} = \sum_k (w_{jk}^{(L)} a_k^{(L-1)}) + b_j^{(L)}$$

# MACHINE LEARNED VIA BACKPROPAGATION

$C_0 \rightarrow$  Cost function

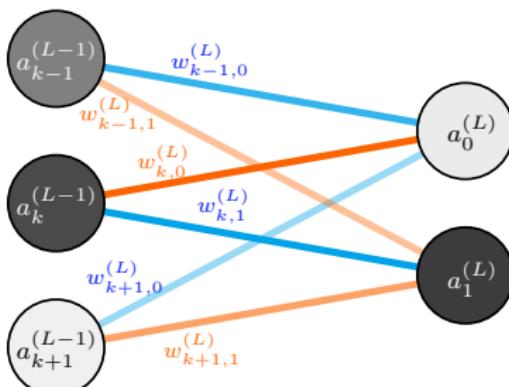
$a_j^{(L)} \rightarrow$  The  $j$ th node's value on layer  $L$

$L \rightarrow$  Layer index

$b_j^{(L)} \rightarrow$  The  $j$ th node's bias on layer  $L$

$y_j \rightarrow$  Desired value of  $a_j^{(L)}$

$w_{jk}^{(L)} \rightarrow$  Weight connecting  $a_k^{(L-1)}$  and  $a_j^{(L)}$



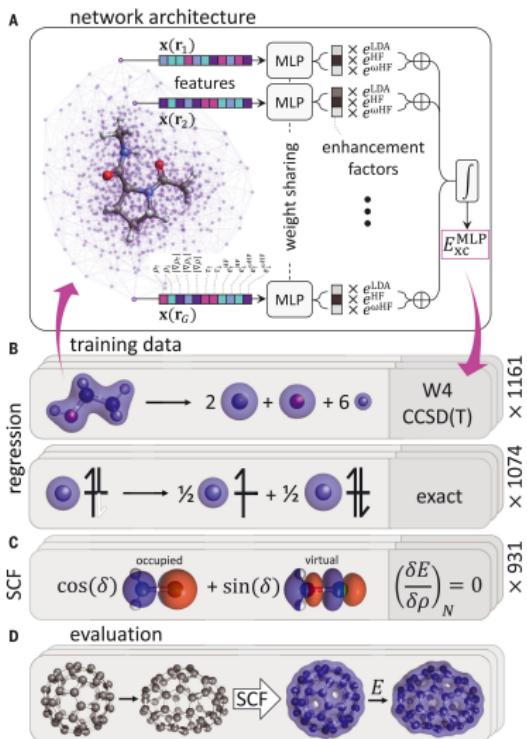
The neural network is “trained” by shifting the biases, weights, and previous layer’s nodes in direction of steepest descent with respect to the cost function.

$$\frac{\partial C_0}{\partial b_j^{(L)}} = \frac{\partial z_j^{(L)}}{\partial b_j^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}}$$

$$\frac{\partial C_0}{\partial w_{jk}^{(L)}} = \frac{\partial z_j^{(L)}}{\partial w_{jk}^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}}$$

$$\frac{\partial C_0}{\partial a_k^{(L-1)}} = \sum_j \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}}$$

# TRAINING DATA (FULL)



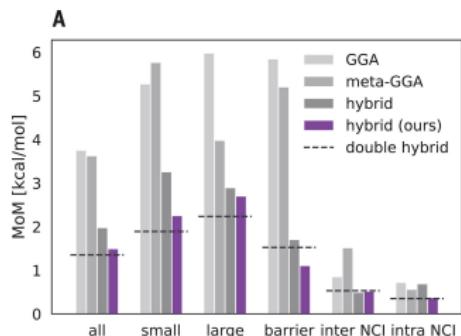
**A)** The selected info at each of the  $G$  atom-centred grid-points form the input feature vectors that are fed into a shared MLP. This is used to predict the  $e_{XC}^{MLP}$  (MLP energy density). Integrating this, and adding the dispersion correction gives the  $E_{XC}^{DM21}$  energy.

**B)** This architecture was trained on 2235 reactions. 1161 represented atomization, ionization, electron affinity, and intermolecular binding energies of small main-group, H-Kr molecules. Another 1074 represented the crucial FC and FS densities only for the atoms H-Ar. Labels were obtained from literature, or CCSD(T).

**C)** For gradient training, perturbation theory gives the leading order change in energy after a single SCF iteration. This energy change depends on the derivatives of the exchange-correlation functional and adding the square of this to the training objective encourages the model to avoid making spuriously large orbital rotations away from reasonable orbitals during self-consistent iteration.

**D)** Once trained, the functional can be deployed in self-consistent calculations.

## BENCHMARKS (FULL)



**B**

|                          | GMTKN55     | BBB          | QM9         |
|--------------------------|-------------|--------------|-------------|
|                          | MoM         |              |             |
| SCAN(:D3 <sub>Bj</sub> ) | 3.63        | 52.90        | 3.57        |
| $\omega$ B97X(-V)        | 2.45        | 70.28        | 2.13        |
| M06-2X                   | 2.08        | 64.50        | 2.20        |
| PW6B95:D3 <sub>0</sub>   | 1.98        | 57.81        | 2.24        |
| DM21                     | <b>1.50</b> | <b>13.20</b> | <b>1.66</b> |

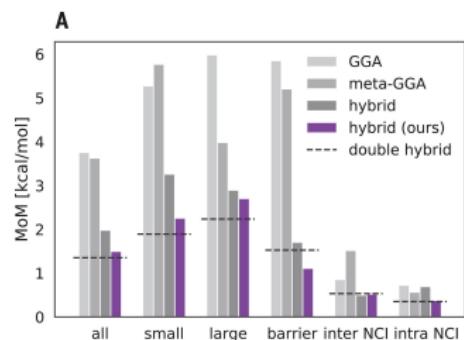
All errors are in kcal/mol, using the mean of means (MoM) error metric.

GMTKN55 1505 reactions across 55 sub-benchmarks of general main group thermochemistry, kinetics and noncovalent interactions, includes heavy atoms beyond Kr.

BBB A new Quantum Monte Carlo bond breaking benchmark: H<sub>2</sub>, Li<sub>2</sub>, C<sub>2</sub>, N<sub>2</sub>, F<sub>2</sub>

QM9 A collection of 133,857 enumerated isomers of organic molecules with up to nine heavy atoms.

## BENCHMARKS (FULL)



**B**

|                          | GMTKN55     | BBB          | QM9         |
|--------------------------|-------------|--------------|-------------|
| MoM                      |             |              |             |
| SCAN(:D3 <sub>BJ</sub> ) | 3.63        | 52.90        | 3.57        |
| $\omega$ B97X(-V)        | 2.45        | 70.28        | 2.13        |
| M06-2X                   | 2.08        | 64.50        | 2.20        |
| PW6B95:D3 <sub>0</sub>   | 1.98        | 57.81        | 2.24        |
| DM21                     | <b>1.50</b> | <b>13.20</b> | <b>1.66</b> |

All errors are in kcal/mol, using the mean of means (MoM) error metric.

**A)** The MoM error metric in each class of reactions from GMTKN55. DM21 is compared with functionals at rungs two to five, with strong GMTKN55 metrics from

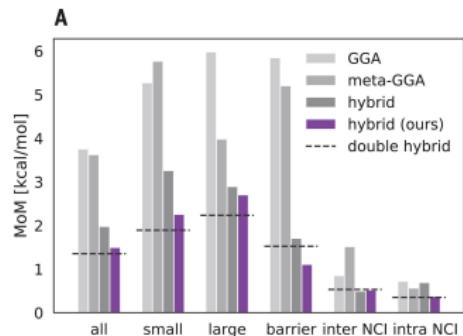
GGA revPBE:D3<sub>BJ</sub>

mGGA SCAN:D3<sub>BJ</sub>,

Hybrid PW6B95:D3<sub>0</sub>

2Hybrid DSD-PBEP86:D3<sub>BJ</sub>

## BENCHMARKS (FULL)



**B**

|                          | GMTKN55     | BBB          | QM9         |
|--------------------------|-------------|--------------|-------------|
|                          | MoM         |              |             |
| SCAN(:D3 <sub>Bj</sub> ) | 3.63        | 52.90        | 3.57        |
| $\omega$ B97X(-V)        | 2.45        | 70.28        | 2.13        |
| M06-2X                   | 2.08        | 64.50        | 2.20        |
| PW6B95:D3 <sub>0</sub>   | 1.98        | 57.81        | 2.24        |
| <b>DM21</b>              | <b>1.50</b> | <b>13.20</b> | <b>1.66</b> |

All errors are in kcal/mol, using the mean of means (MoM) error metric.

**B)** Performance of DM21 compared with the SCAN functional and the three best performing hybrid functionals on three benchmark sets.

- The BBB benchmark measures mean absolute errors for selected first- and second-row diatomics relative to high-level UCCSD(T) (cation) and QMC (neutral) calculations.
- QM9 errors are taken relative to high-level G4(MP2) theory.
- “:D3<sub>Bj</sub>” indicates that the best of SCAN with or without D3 correction is reported on each metric and similarly for VV10 variants of  $\omega$ B97X.

## BASIS & CODE

They mostly used Karlsruhe (aug'-)def2-QZVP basis sets throughout. Specifically in their validation and in their benchmarks.

aug Augmented versions of the preceding basis sets with added diffuse functions

def2 newer redefinition

QZVP Valence quadruple-zeta polarization (applies for H-La and Hf-Rn)

The pySCF interface for DM21 can be acquired via their GitHub:

[https://github.com/deepmind/deepmind-research/tree/master/density\\_functional\\_approximation\\_dm21](https://github.com/deepmind/deepmind-research/tree/master/density_functional_approximation_dm21)

Included is detailed instructions and information regarding

- Installing directly
- Downloading and installing from a local git repository
- PySCF interface
- Best practices for using the neural functionals.
- Using DM21 from C++
- Detailed explanation