



The
University
Of
Sheffield.

Machine Learning in Chemical Simulation: Solids, Liquids and Biochemistry

Chris Handley

Department of Materials Science and Engineering
c.m.handley@sheffield.ac.uk

*The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, **and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed,** which can lead to an explanation of the main features of complex atomic systems without too much computation.*

Paul Dirac - Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, Vol. 123, No. 792 (6 April 1929)

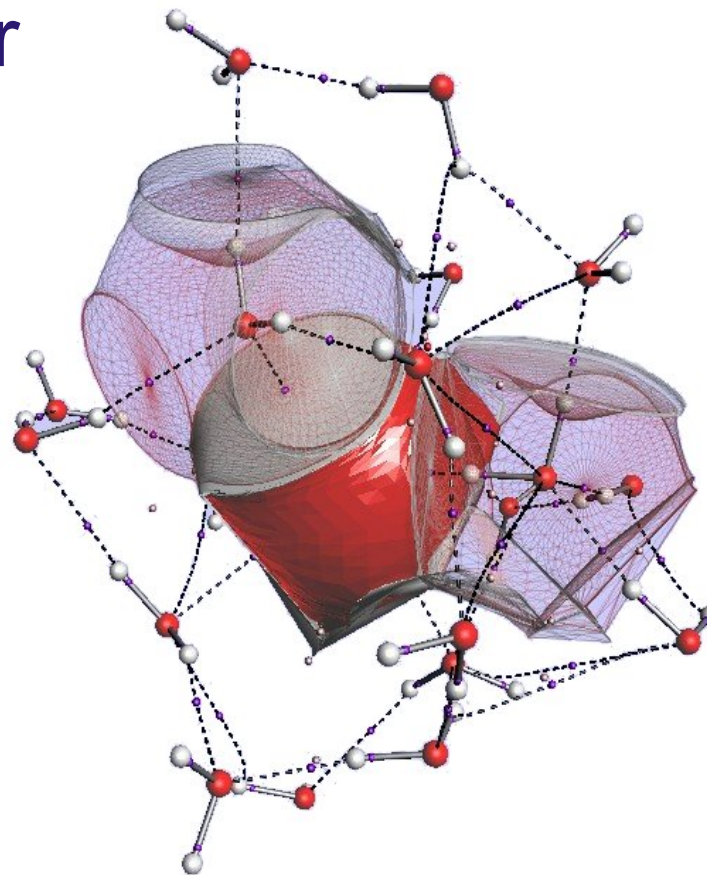
An increase in computer power of at least two orders of magnitude should occur over the next decade. *Without further research into the accuracy of force-field potentials, future macromolecular modeling may well be limited more by validity of the energy functions, particularly electrostatic terms, than by technical ability to perform the computations.*

Force fields for protein simulations

JW Ponder, DA Case - Advances in protein chemistry, 2003

Popelier Group, Manchester – UMIST and University of Manchester

- Electron density – 3D regions
- Mapped by QCT
- Represented by spherical harmonics

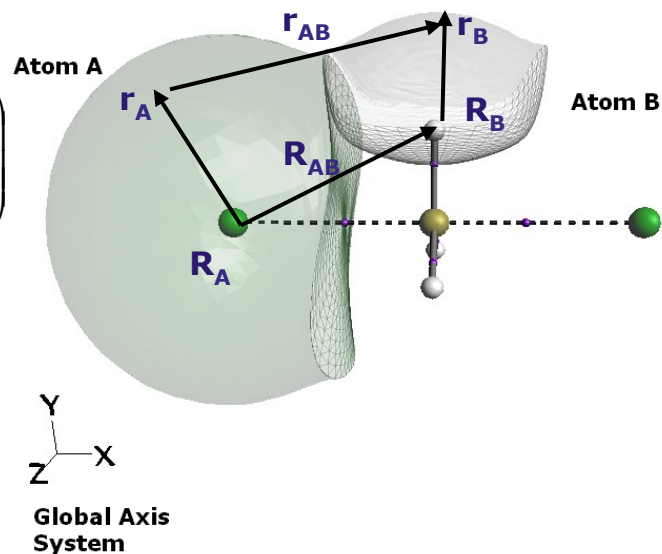


Popelier Group, Manchester – UMIST and University of Manchester

$$E(\mathbf{r}^N) = \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 +$$

$$\sum_{\text{torsions}} \frac{V_n}{2} (1 + \cos(n\varpi - \gamma)) + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)$$

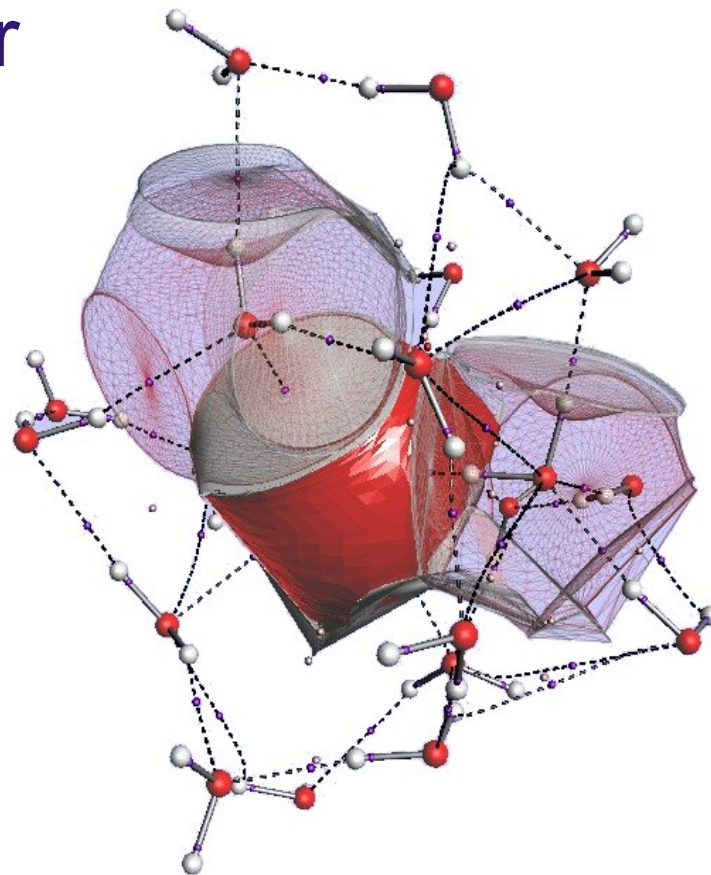
$$E_{\text{elec}}(A, B) = \int d\mathbf{r}_A d\mathbf{r}_B \frac{\rho(\mathbf{r}_A)\rho(\mathbf{r}_B)}{|\mathbf{R}_{AB} - (\mathbf{r}_A - \mathbf{r}_B)|}$$



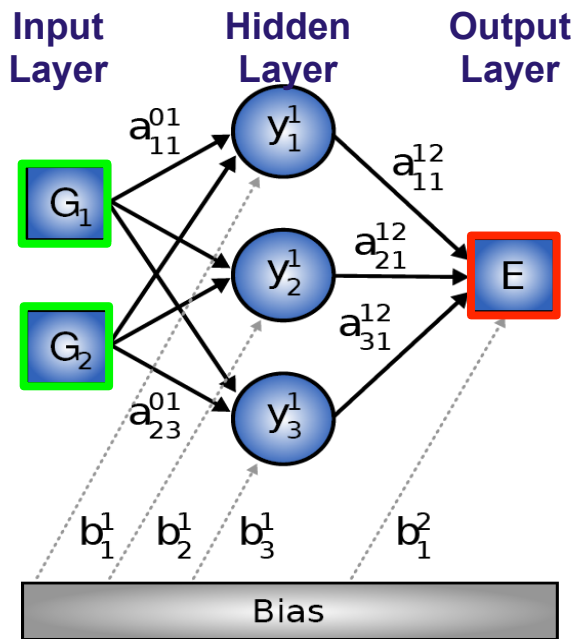
$$E_{\text{elec}}(A, B) = \sum_{l_A=0}^{\infty} \sum_{l_B=0}^{\infty} \sum_{m_A=-l_A}^{l_A} \sum_{m_B=-l_B}^{l_B} T_{l_A m_A l_B m_B}(\mathbf{R}) Q_{l_A m_A}(\Omega_A) Q_{l_B m_B}(\Omega_B)$$

Popelier Group, Manchester – UMIST and University of Manchester

- Spherical harmonics depend on local chemical environment.
- Sample as many configurations for each atom.



Popelier Group, Manchester – UMIST and University of Manchester



Analytic Expression:

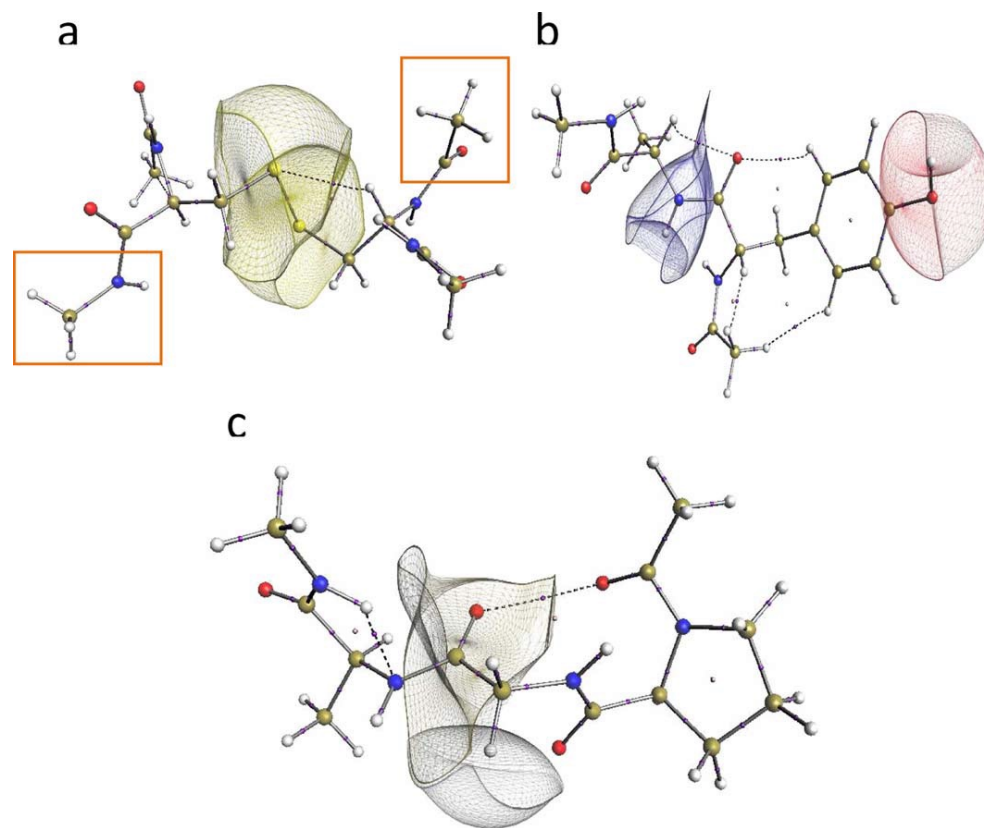
$$E = f_1^2 \left(b_1^2 + \sum_{j=1}^3 a_{j1}^{12} \cdot f_j^1 \left(b_j^1 + \sum_{i=1}^2 G_i \cdot a_{ij}^{01} \right) \right)$$

$$\{G\} \xrightarrow{NN} E$$

Structure Energy

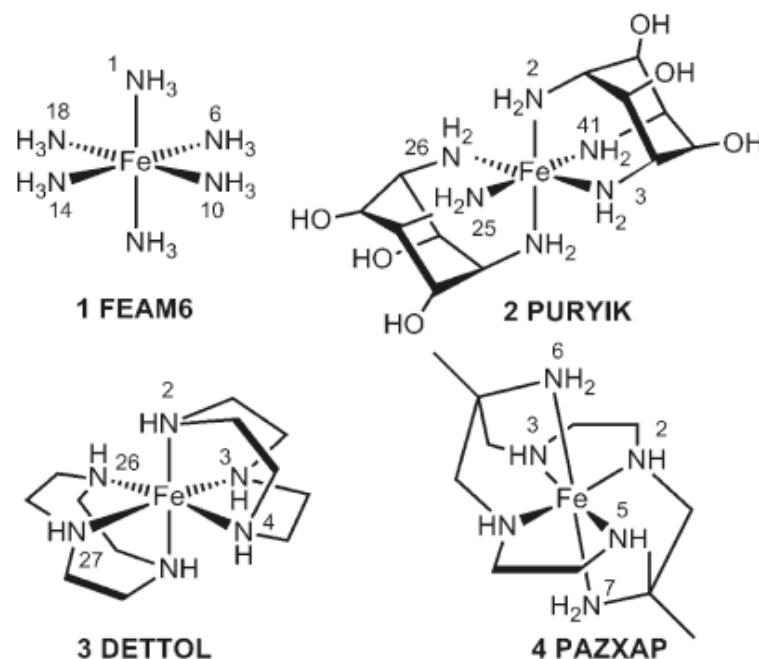
Popelier Group, Manchester – UMIST and University of Manchester

- PhD – Polarizable Multipolar Peptides and Polarizable Ions in Water.
- This work formed the basis for the ongoing work on a novel force field.
- Some forays into using other machine learning methods. Kriging GPRs



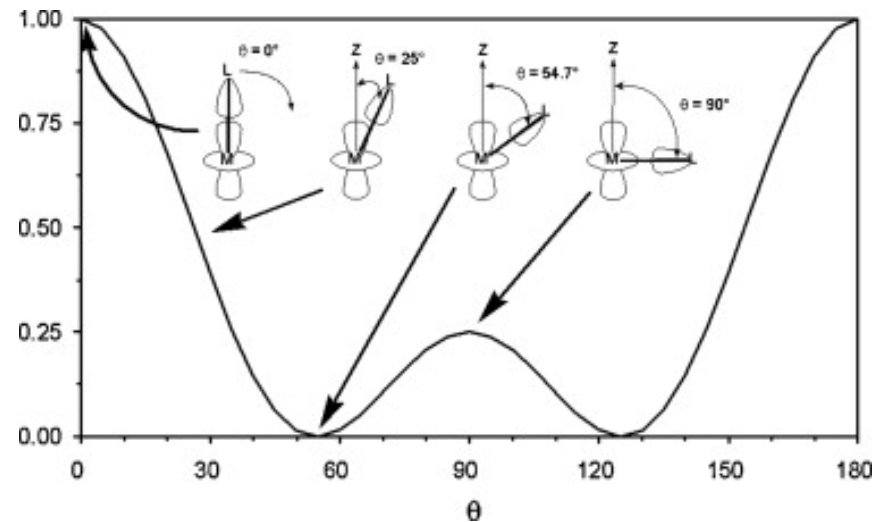
Deeth Group, University of Warwick

- **Ligand Field Molecular Mechanics**
 - d-electrons explicitly represented in force fields.
 - Vaguely similar to Tight Binding.
- Implemented in the CCG MOE suite.



Deeth Group, University of Warwick

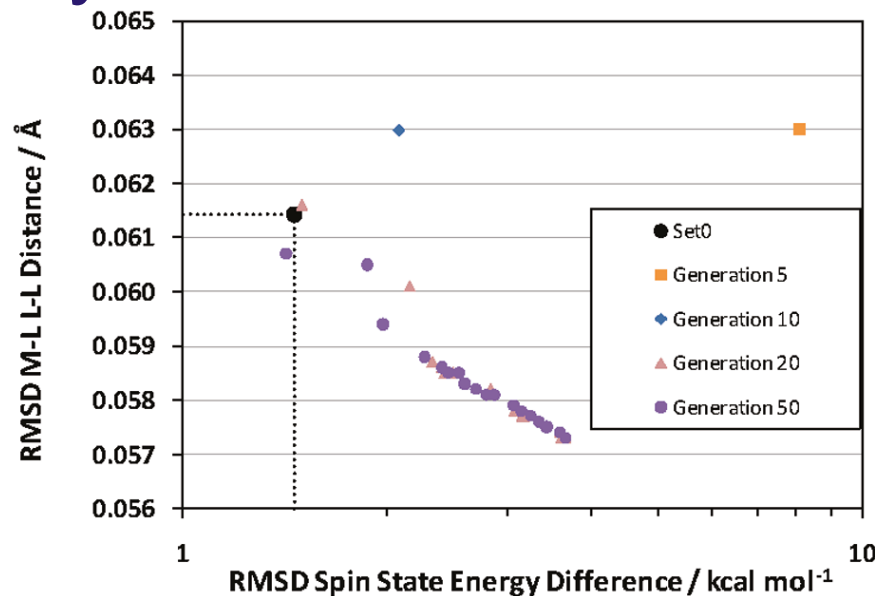
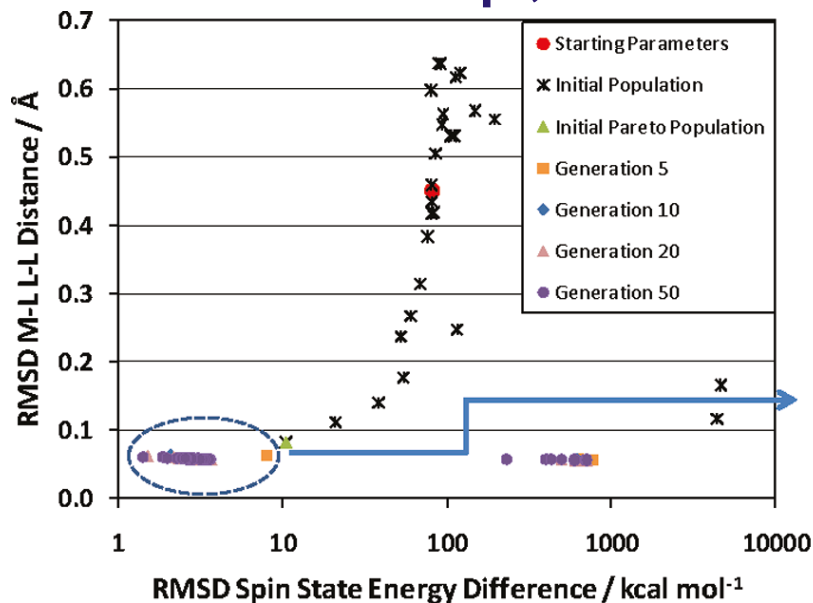
- Method introduces new functions – LFMM just for this one system type introduces 30+ new parameters.
- By hand with a priori expert knowledge fitting by hand can take 6 months.



$$E(d_{z^2}) = F_{\sigma}^2(d_{z^2})e_{\sigma} = \frac{1}{16}(1 + 3 \cos 2\theta)^2 e_{\sigma}$$

$$\langle d_i | V_{LF} | d_j \rangle = \sum_l \sum_k^{N \text{ symm}} F_{ik}^l F_{kj}^l e_k^l$$

Deeth Group, University of Warwick

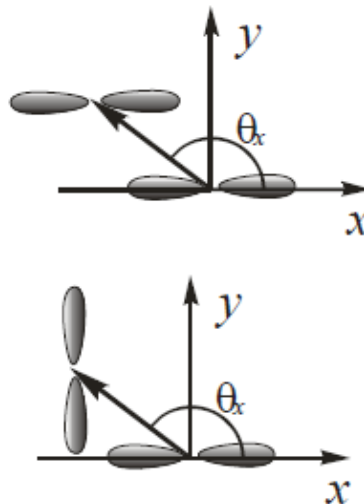


- **Multi-Objective Evolutionary Algorithms** – optimize using Genetic Algorithms, but guided by Pareto Front analysis.
- Replicate fitting, and improve on fitting, in an hour!
- First example of method implemented for force field design.

Behler Group, Ruhr-Universität Bochum, Germany

- Neural Network approach to Tight Binding.

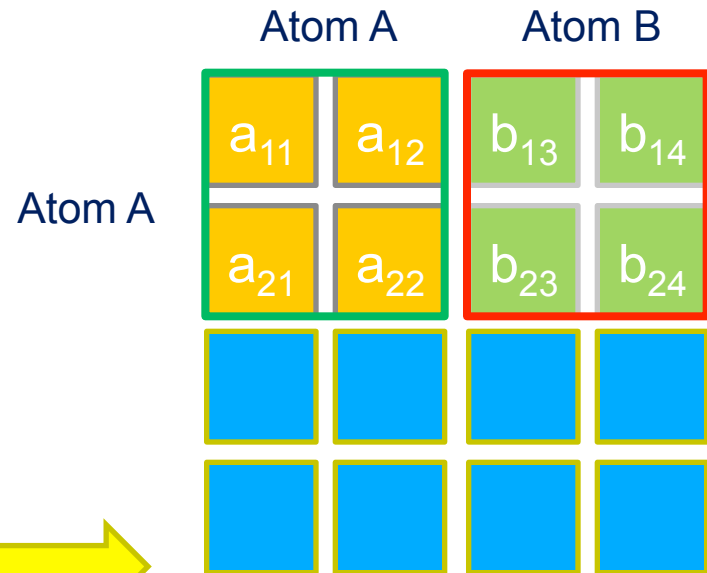
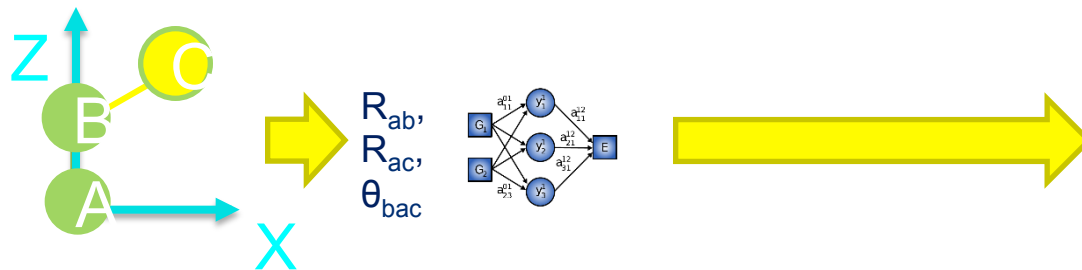
$$\begin{aligned}
 h_{p_x p_x} &= s(r) [l^2 V_{pp\sigma} + (1 - l^2 V_{pp\pi})] \\
 l &= \cos(\theta_x)
 \end{aligned}$$



$$\begin{aligned}
 E_{s,s} &= V_{ss\sigma} \\
 E_{s,x} &= lV_{sp\sigma} \\
 E_{x,x} &= l^2 V_{pp\sigma} + (1 - l^2) V_{pp\pi} \\
 E_{x,y} &= lmV_{pp\sigma} - lmV_{pp\pi} \\
 E_{x,z} &= lnV_{pp\sigma} - lnV_{pp\pi} \\
 E_{s,xy} &= 3^{1/2} lmV_{sd\sigma} \\
 E_{s,x^2-y^2} &= \frac{1}{2} 3^{1/2} (l^2 - m^2) V_{sd\sigma} \\
 E_{s,3z^2-r^2} &= [n^2 - \frac{1}{2}(l^2 + m^2)] V_{sd\sigma} \\
 E_{x,xy} &= 3^{1/2} l^2 mV_{pd\sigma} + m(1 - 2l^2) V_{pd\pi} \\
 E_{x,yz} &= 3^{1/2} lmnV_{pd\sigma} - 2lmnV_{pd\pi} \\
 E_{x,zx} &= 3^{1/2} l^2 nV_{pd\sigma} + n(1 - 2l^2) V_{pd\pi} \\
 E_{x,x^2-y^2} &= \frac{1}{2} 3^{1/2} l(l^2 - m^2) V_{pd\sigma} + l(1 - l^2 + m^2) V_{pd\pi} \\
 E_{y,x^2-y^2} &= \frac{1}{2} 3^{1/2} m(l^2 - m^2) V_{pd\sigma} - m(1 + l^2 - m^2) V_{pd\pi} \\
 E_{z,x^2-y^2} &= \frac{1}{2} 3^{1/2} n(l^2 - m^2) V_{pd\sigma} - n(l^2 - m^2) V_{pd\pi} \\
 E_{x,3z^2-r^2} &= l[n^2 - \frac{1}{2}(l^2 + m^2)] V_{pd\sigma} - 3^{1/2} l n^2 V_{pd\pi} \\
 E_{y,3z^2-r^2} &= m[n^2 - \frac{1}{2}(l^2 + m^2)] V_{pd\sigma} - 3^{1/2} m n^2 V_{pd\pi} \\
 E_{z,3z^2-r^2} &= n[n^2 - \frac{1}{2}(l^2 + m^2)] V_{pd\sigma} + 3^{1/2} n(l^2 + m^2) V_{pd\pi} \\
 E_{xy,xy} &= 3l^2 m^2 V_{dd\sigma} + (l^2 + m^2 - 4l^2 m^2) V_{dd\pi} + (n^2 + l^2 m^2) V_{dd\delta} \\
 E_{xy,yz} &= 3lm^2 nV_{dd\sigma} + ln(1 - 4m^2) V_{dd\pi} + ln(m^2 - 1) V_{dd\delta} \\
 E_{xy,zx} &= 3l^2 mnV_{dd\sigma} + mn(1 - 4l^2) V_{dd\pi} + mn(l^2 - 1) V_{dd\delta} \\
 E_{xy,x^2-y^2} &= \frac{3}{2} lm(l^2 - m^2) V_{dd\sigma} + 2lm(m^2 - l^2) V_{dd\pi} + \frac{1}{2} lm(l^2 - m^2) V_{dd\delta} \\
 E_{yz,x^2-y^2} &= \frac{3}{2} mn(l^2 - m^2) V_{dd\sigma} - mn[1 + 2(l^2 - m^2)] V_{dd\pi} \\
 &\quad + mn[1 + \frac{1}{2}(l^2 - m^2)] V_{dd\delta} \\
 E_{zx,x^2-y^2} &= \frac{3}{2} n(l^2 - m^2) V_{dd\sigma} + n/[1 - 2(l^2 - m^2)] V_{dd\pi} \\
 &\quad - n/[1 - \frac{1}{2}(l^2 - m^2)] V_{dd\delta} \\
 E_{xy,3z^2-r^2} &= 3^{1/2} lm[n^2 - \frac{1}{2}(l^2 + m^2)] V_{dd\sigma} - 3^{1/2} 2lmn^2 V_{dd\pi} \\
 &\quad + \frac{1}{2} 3^{1/2} lm(1 + n^2) V_{dd\delta} \\
 E_{yz,3z^2-r^2} &= 3^{1/2} mn[n^2 - \frac{1}{2}(l^2 + m^2)] V_{dd\sigma} + 3^{1/2} mn(l^2 + m^2 - n^2) V_{dd\delta}
 \end{aligned}$$

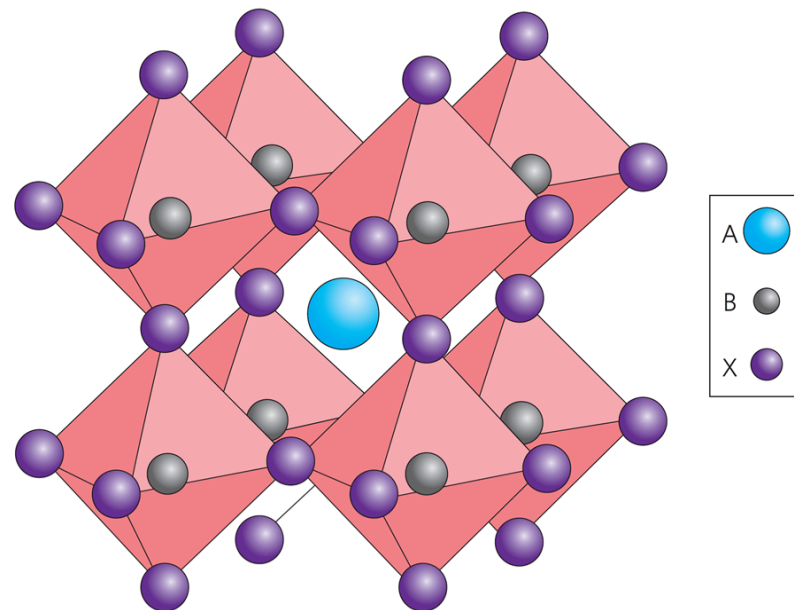
Behler Group, Ruhr-Universität Bochum, Germany

$$\begin{aligned} & \langle i\mu | \hat{V}_{ext} | j\nu \rangle \\ &= \langle i\mu | \hat{V}_{ext}^{(i)} | j\nu \rangle + \langle i\mu | \hat{V}_{ext}^{(j)} | j\nu \rangle \\ &+ \sum_{k \neq i,j} \langle i\mu | \hat{V}_{ext}^{(k)} | j\nu \rangle \end{aligned}$$



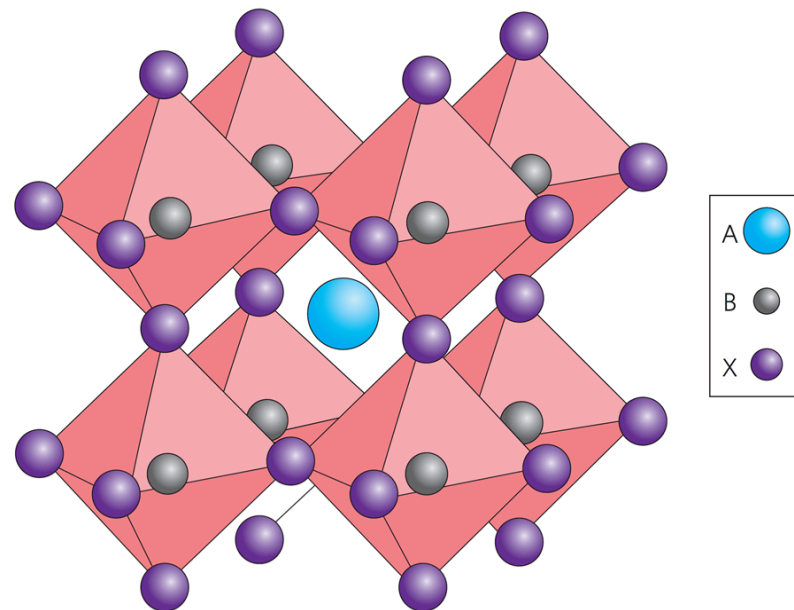
Reaney, Sinclair, Freeman, Dean, University of Sheffield

- Perovskites – DFT simulations dominate literature. Force Field methods could reveal more about the dynamics of the structures.
- A need to develop new potentials that capture the right chemistry.



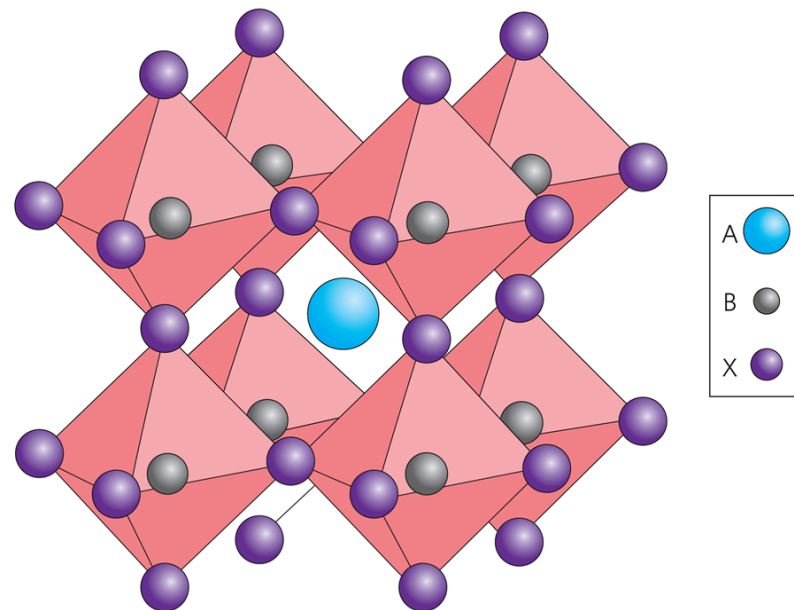
Reaney, Sinclair, Freeman, Dean, University of Sheffield

- Multi-layer capacitors – we know the materials but how do we combine them?
- A machine learning problem – permutation of combinations of layers, thickness of layers etc.
- Develop computational tools for synthetic chemists.



Reaney, Sinclair, Freeman, Dean, University of Sheffield

- Identify places where chemical simulation can support and guide synthesis.
- Identify places where previous machine learning methods can aid design of materials i.e. force field design or materials classification.





The
University
Of
Sheffield.

Questions?