

Machine Learning Surrogate Integration for Efficient Computer-Aided Molecular and Process Design

Florian Baakes, George Jackson, Amparo Galindo, Claire S. Adjiman

Department of Chemical Engineering, Sargent Centre of Process Systems Engineering, Institute for Molecular Science and Engineering, Imperial College London, SW7 2AZ, London, UK

Motivation¹

- Conventional CO₂ capture imposes a **20–30%** parasitic energy **penalty**
- Enormous chemical design space** (~10⁶³ candidate molecules)
- Computer-aided **molecular** and **process design** (CAMPD) guides **discovery**

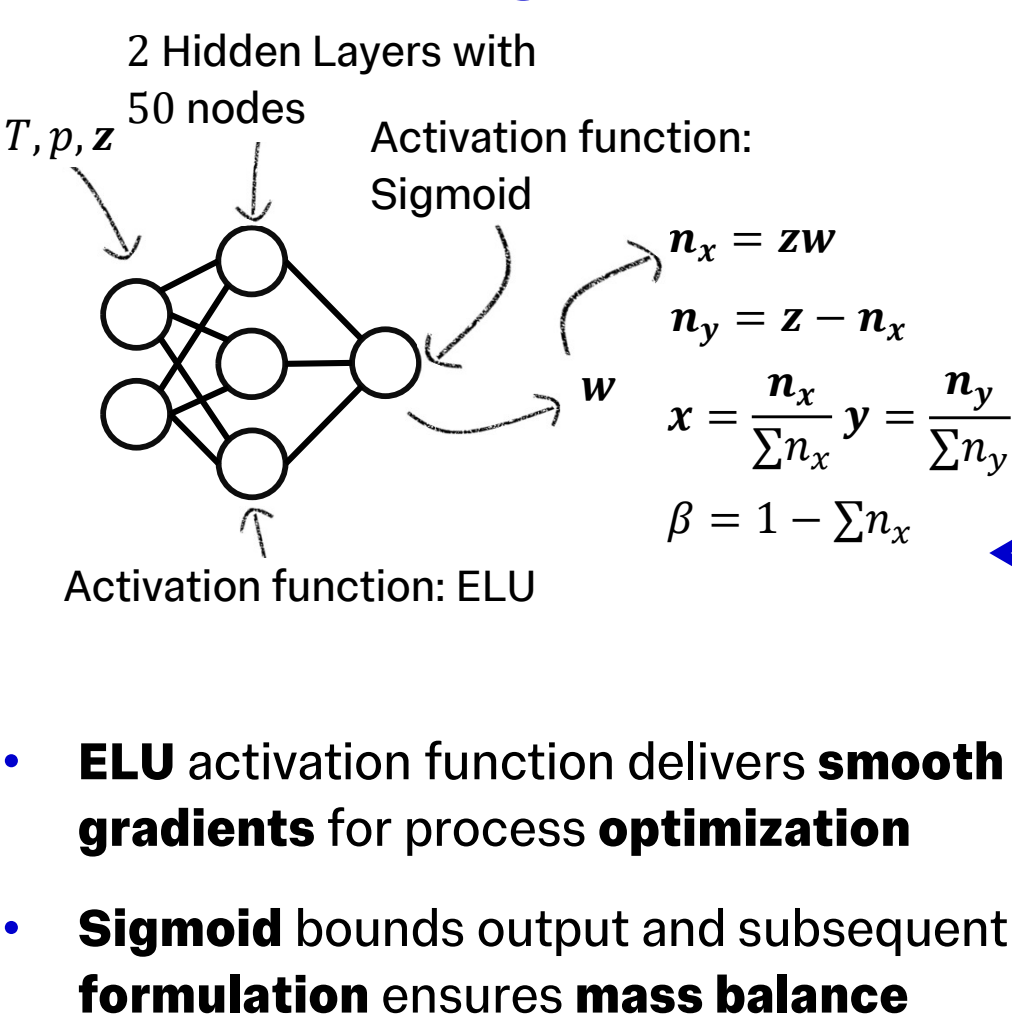
Challenges

- Expensive *Tp*-flash** calculations **dominate** optimization **times**
- Surrogate** models speed up optimizations but **need large datasets**
- Transfer-learning** reduces this burden, yielding a **net computational speed-up**

Training^{6,7}

- Initialized** from **Monoethanolamine (MEA)**/ **2,2-bis(aminomethyl)propan-1,3-diol (BAMPD)** pre-trained models
- Single seed** vs. greedy soup averaging of **9 seeds**
- Adaptive** learning rate: 0.01 – 0.00005
- 250 epochs** → **20, 50, 110 s** training time

ANN Surrogate^{6,7}



Training data

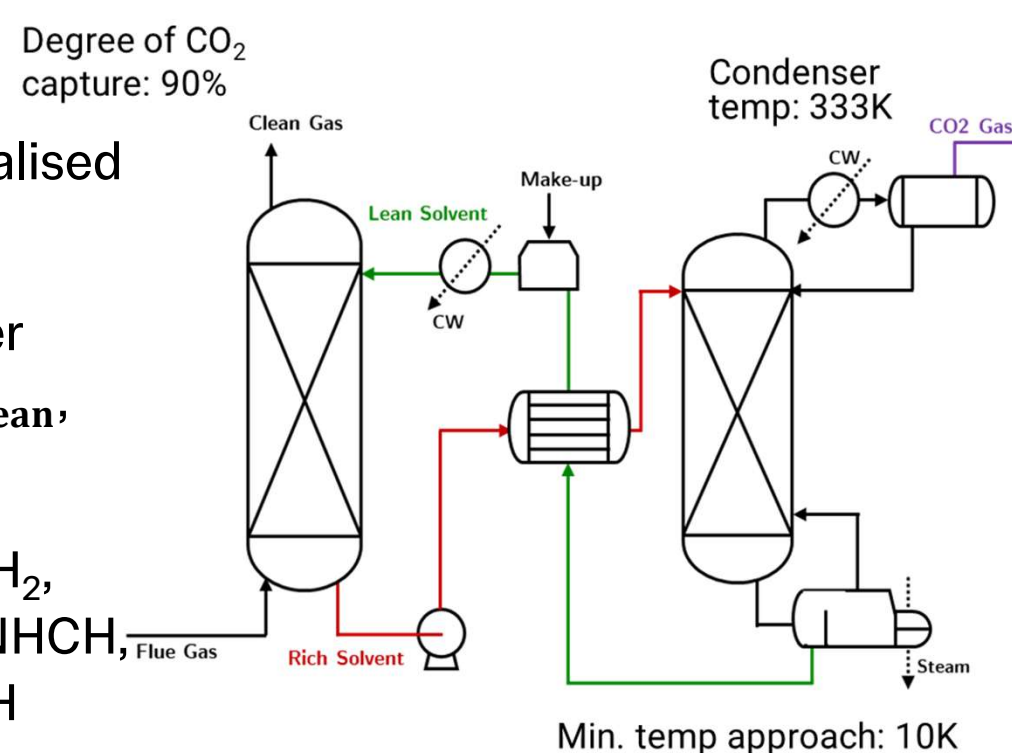
- Chemical and phase equilibria** of quaternary mixture sampled using **Sobol'** sequence:
- $T \in [298.15, 410]$ K,
 $p \in [0.095, 0.35]$ MPa,
 $z_{H_2O} \in [0.48, 0.984]$,
 $z_{Amine} \in [0.001, 0.15]$,
 $z_{CO_2} \in [0.005, 0.25]$,
 $z_{N_2} \in [0.0, 0.35]$
- 512, 1024, 2048** data points → **8, 15, 30 s** sampling time
- Z8 HP workstation, 32 GB working memory, Intel Xeon 5416S processor, NVIDIA T1000 graphics card.

Thermodynamic model²⁻⁵

- SAFT-γMie** group contribution **equation of state**
- Represents **molecules** as spherical **groups** and combines them to form **new molecules (like Lego™)**
- Only equation of state **capable of predicting CO₂ absorption** for solvents **never before synthesized**

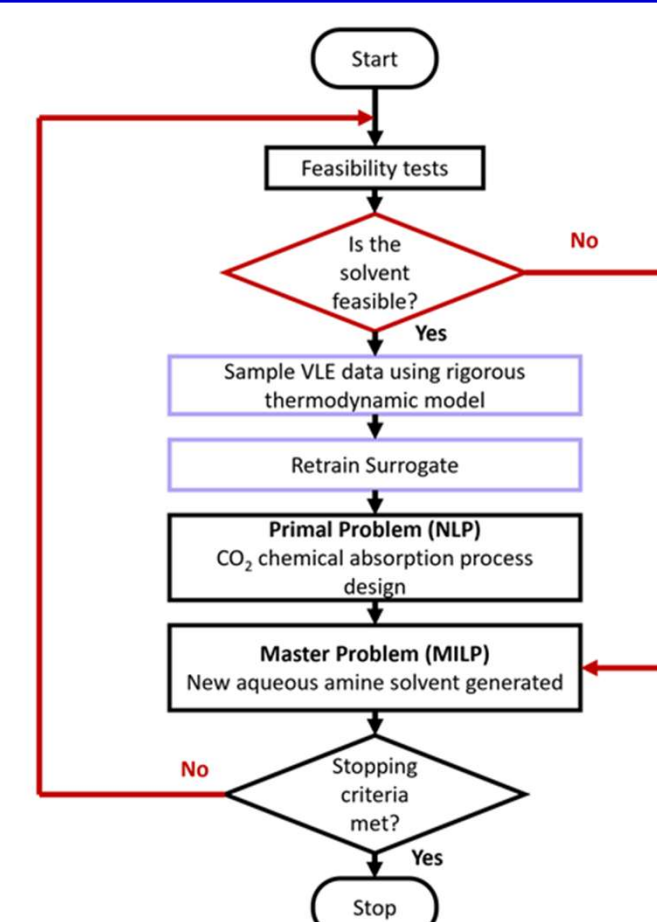
Case study⁸

- 400MWe** CCGT Power Plant
- Objective:** Minimise Total Annualised Cost (**TAC**)
- Process design** space: Desorber pressure p_{des} , Lean loading θ_{lean} , Absorber temperature T_{Abs}
- Molecular design** space: NH₂CH₂, NH₂CH, NH₂C, NHCH₃, NHCH₂, NHCH, NCH₃, NCH₂, CH₃, CH₂, CH, C, OH



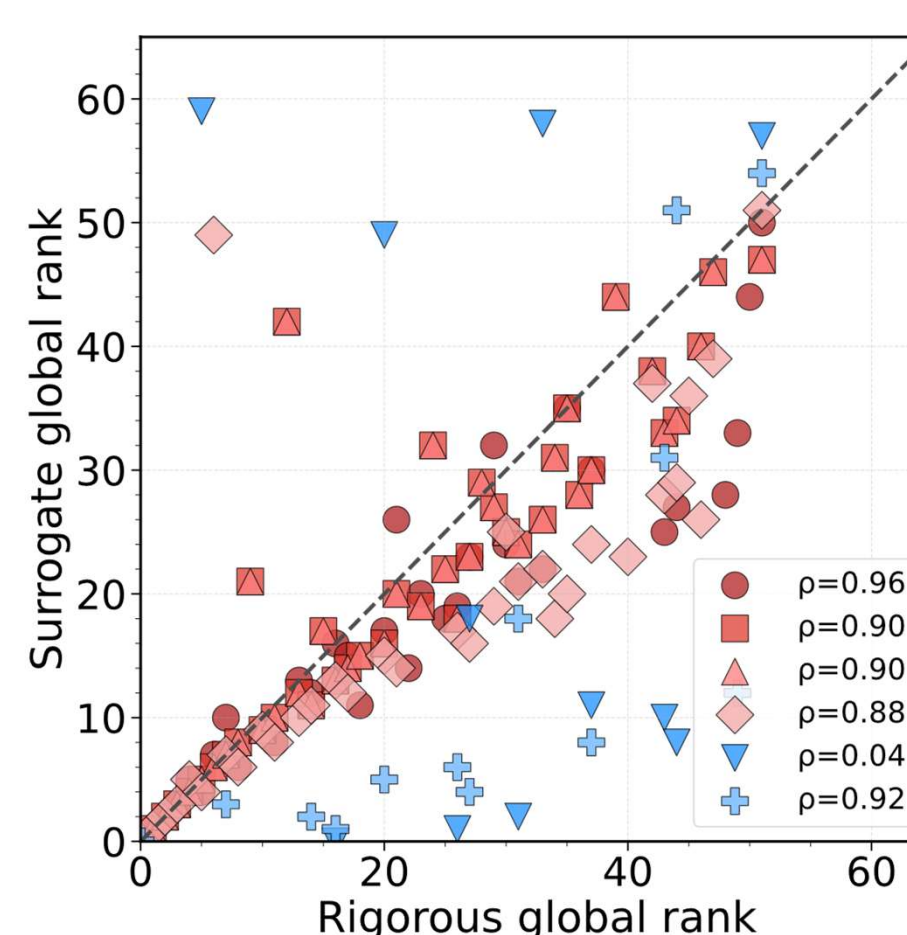
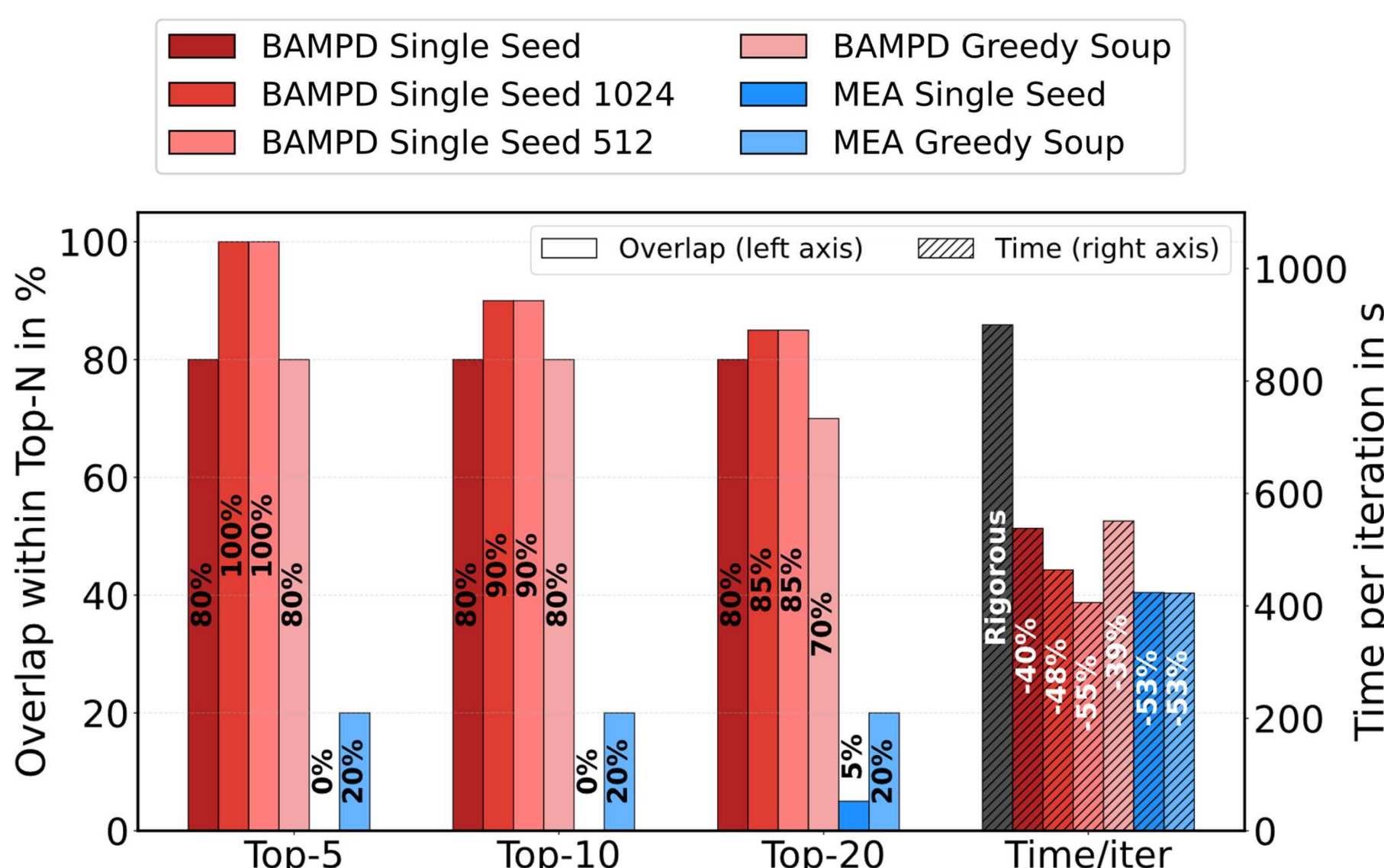
A robust solution strategy⁸

- Feasibility tests** added to outer approximation method for **robust** solution of MINLP
- Test 1: Solvent **property feasibility**, Test 2: Solvent **miscibility**, Test 3: **Absorption capacity**
- Transfer-learned **surrogates** replace **costly *Tp*-flash calculations** during NLP optimization



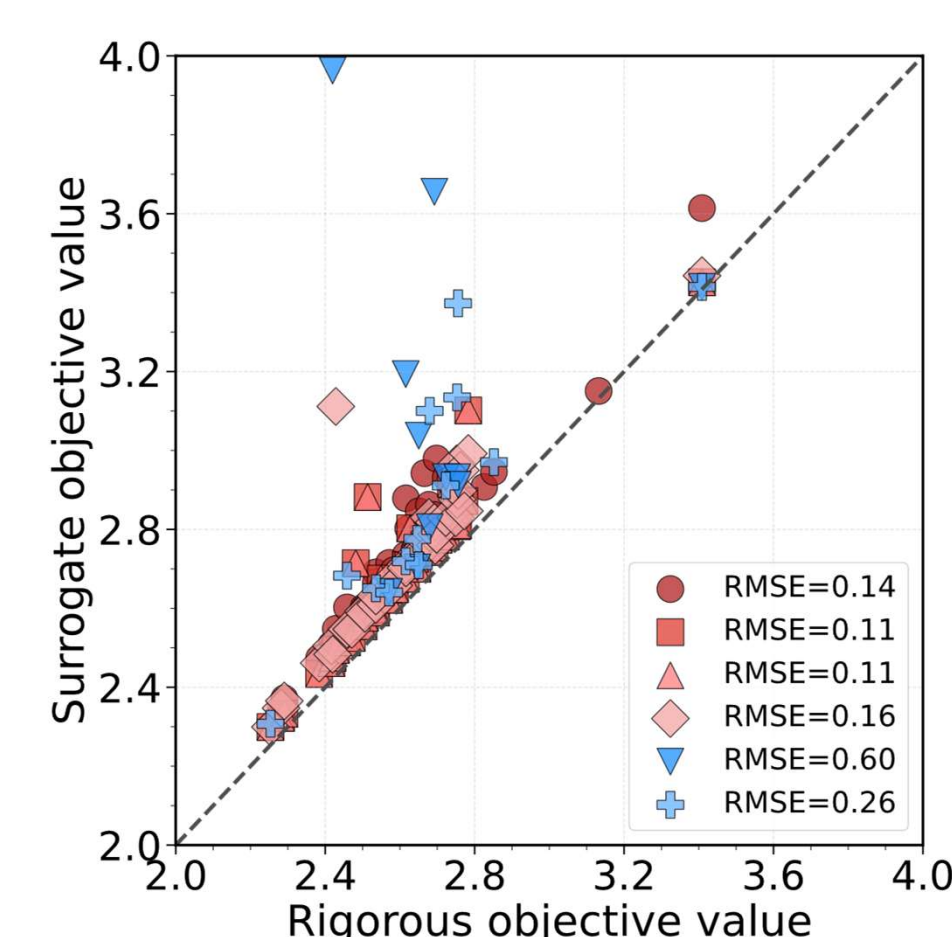
Results

- Base models:** initialized from **MEA** or **BAMPD**; **single-seed** vs. **greedy-soup**
- Best model repeated** with **1024** and **512** data points



- Rank order** mostly recovered with most **Shearman** coefficient close to **0.9**
- Greedy soup** averaging **worsens** results for well-performing **models**

- Objective** function values show **similar trend** to rigorous model
- Slight **offset** due to **insensitivity** with respect to desorber **pressure**



Conclusion

- Transfer-learned **surrogates** reduce CAMPD iteration **time** from **900 s** → **400 s** while **recovering 90 % of Top-10** molecules, Spearman $\rho = 0.96$ vs. rigorous model
- Selecting a **suitable base model** is **paramount** for stable **transfer behaviour**
- Future studies: **multi-start optimization** amplifying **time reduction** benefits