NICK SAHINIDIS' RESEARCH OVERVIEW

OPTIMIZATION ALGORITHMS

- BARON
  - Convexification
  - Branching
- Derivative-free optimization
- GPU algorithms
  - Least squares

MACHINE LEARNING

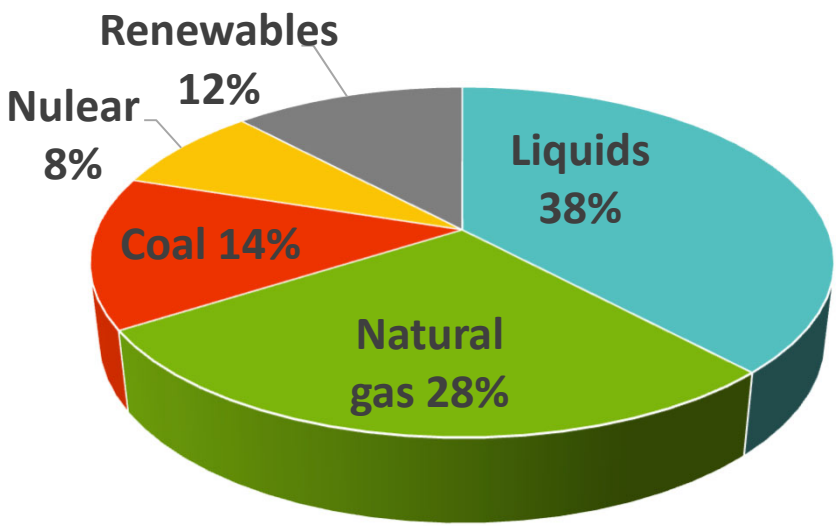
- ALAMO
  - Best subset selection
  - Symbolic regression
  - Constrained regression

MODELING & APPLICATIONS

- Enterprise-wide optimization
  - Maintenance scheduling
  - Uncertainty
- Thermodynamics
  - New equations of state
- Kinetics
- Text mining
  - Maintenance records
  - Patent ranking
- Molecular design
  - Coolants for electronic equipment

THE ENERGY PROBLEM

- **Energy consumption**
  - **Today: 15-20 TW**
  - **Will double by 2040-2050**
  - **We have plenty of energy sources**

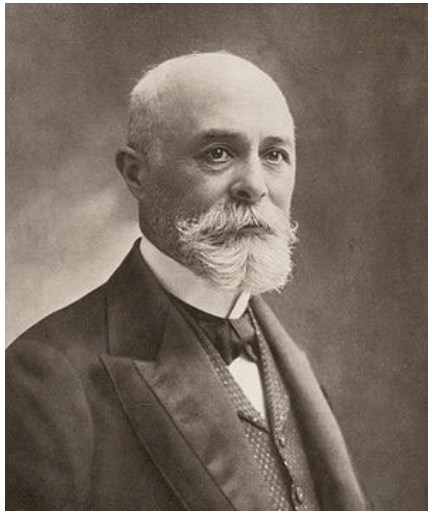


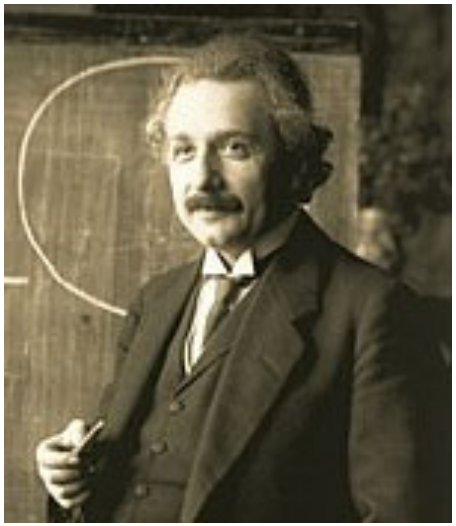
- **Concerns about CO_2**
- **Nuclear, biomass, hydro, geothermal, tide, wind**
  - **Insufficient or impractical**
- **Solar: 10000 TW**
  - **Solar thermal**
  - **Solar electric**
  - **Solar fuel**

www.iea.org

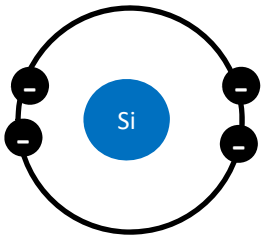
HISTORY OF PHOTOVOLTAICS

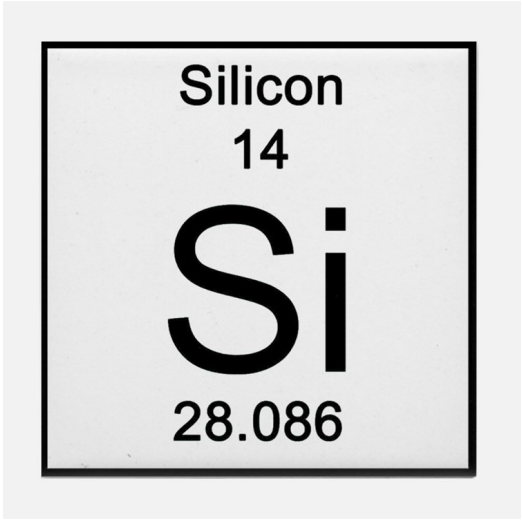
- **Becquerel**
  - **Photovoltaic effect (1839)**
  - **Creation of voltage and current in a material upon exposure to light**
- **Hertz**
  - **Observation of photoelectric effect (1887)**
  - **Emission of electrons when light shines on a material**
- **Einstein**
  - **Explanation of photoelectric effect (1921)**



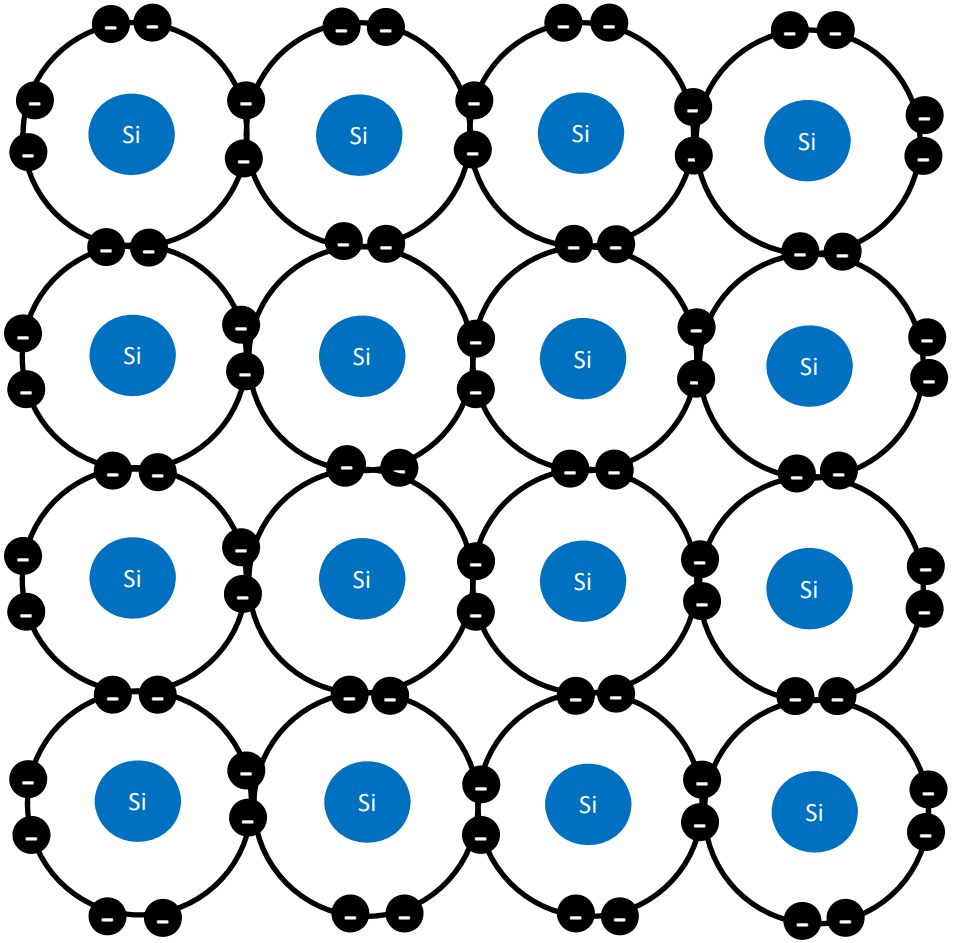


HOW SILICON SOLAR CELLS WORK

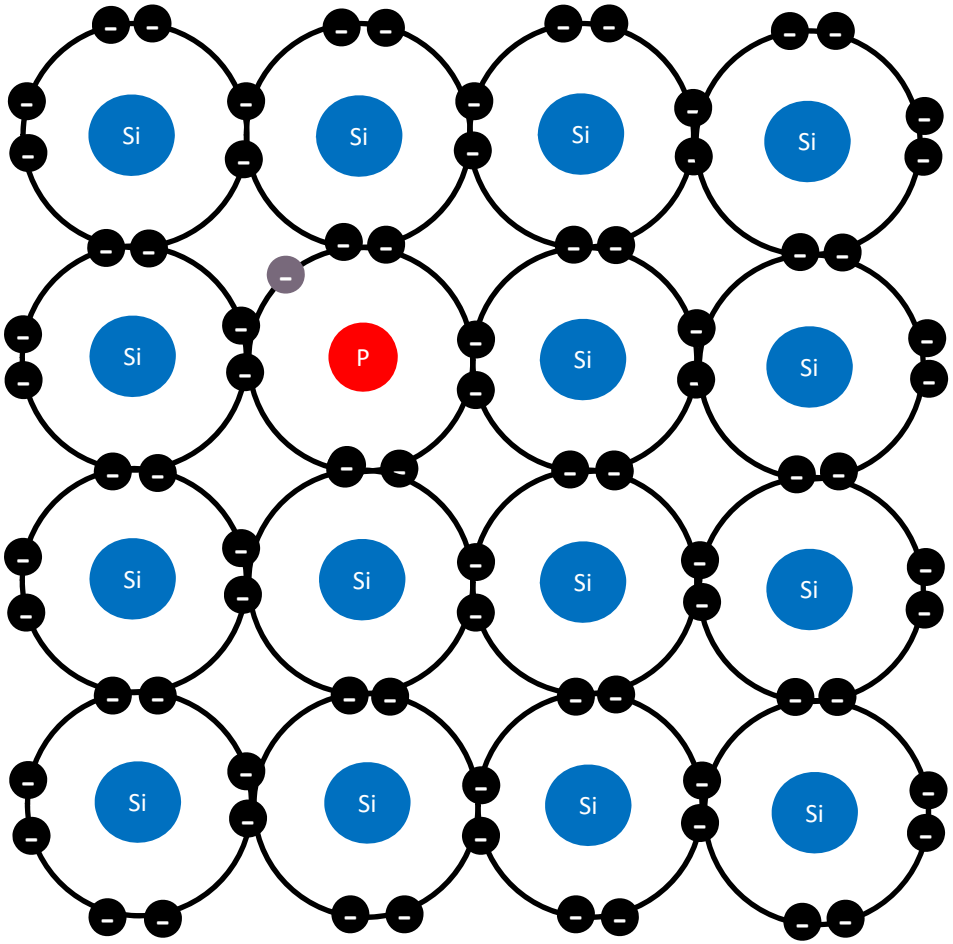




HOW SILICON SOLAR CELLS WORK

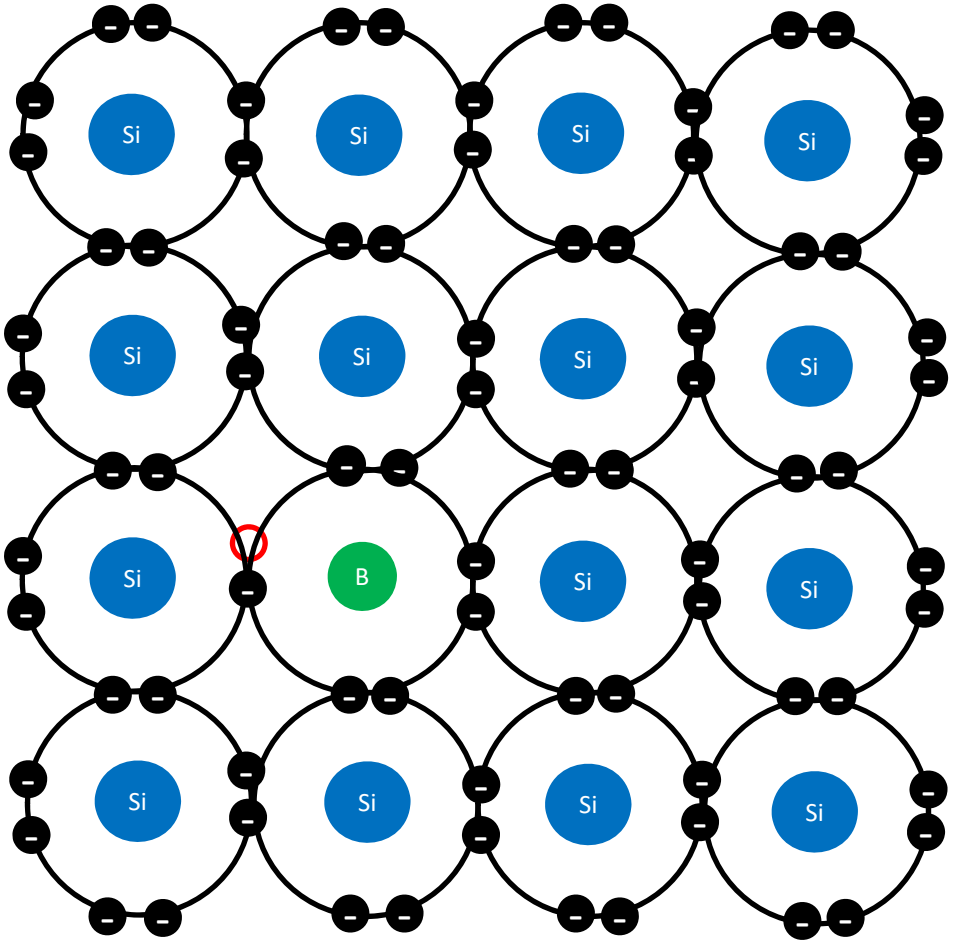


HOW SILICON SOLAR CELLS WORK



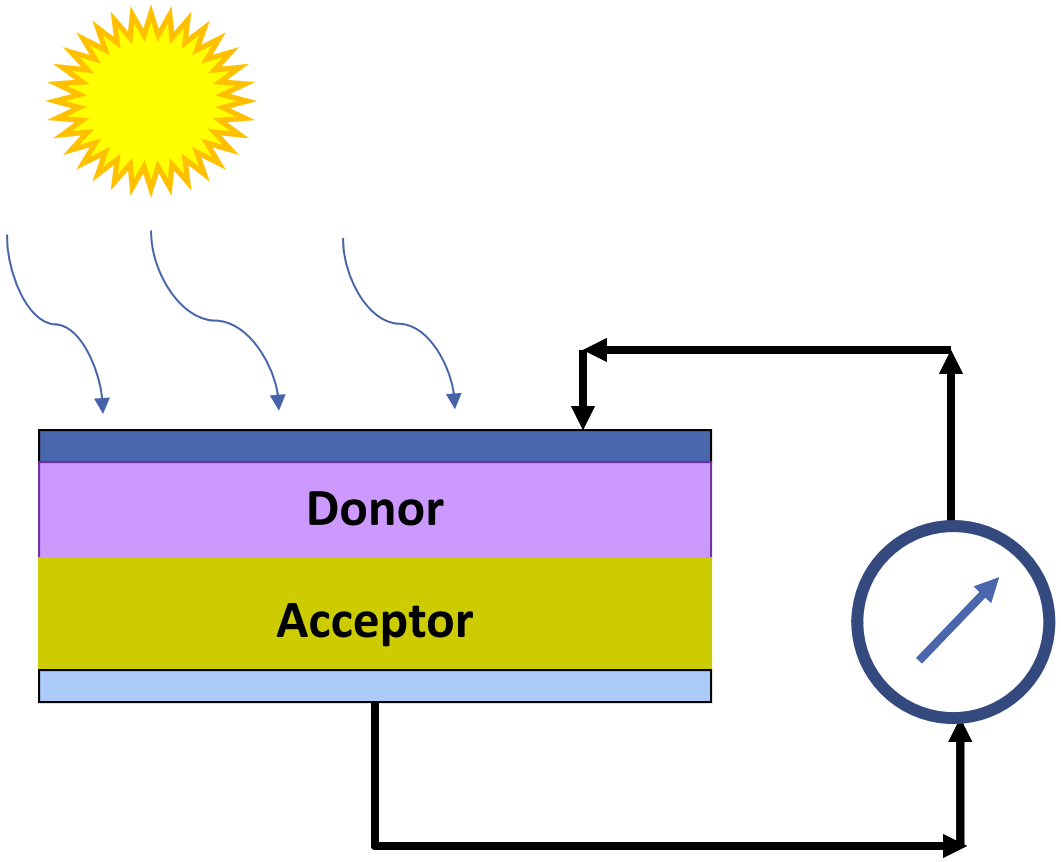
N-type semiconductor (donor)

HOW SILICON SOLAR CELLS WORK



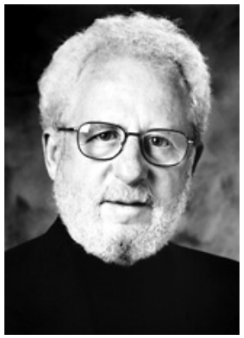
P-type semiconductor (acceptor)

HOW SILICON SOLAR CELLS WORK



THREE GENERATIONS OF SOLAR CELLS

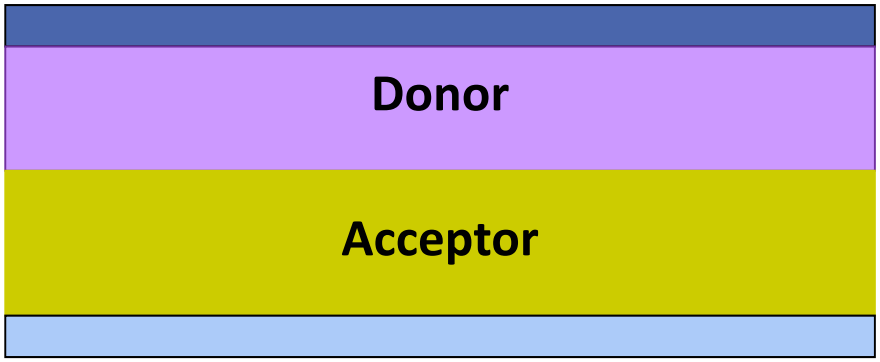
1. **Crystalline silicon**
   - Bell labs (1950); 5% efficiency
   - Expensive; made in high temperatures; high material usage
2. **Thin firm solar cells**
   - Lower temperatures, lower material usage
   - Instability (e.g., water/oxygen)
3. **Third generation**
   - Thin, low Ts, fast to produce etc.
   - Discovery and development of conductive polymers
     - ***Heeger, MacDiarmid, Shirakawa (2000)***

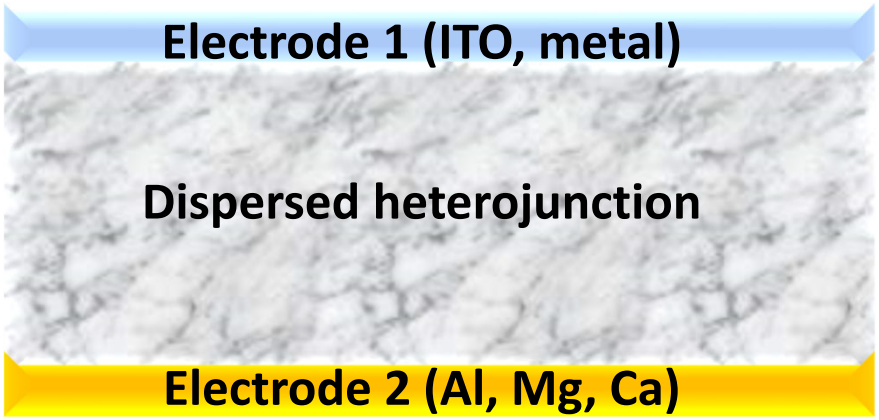

Alan J. Heeger


Alan G. MacDiarmid

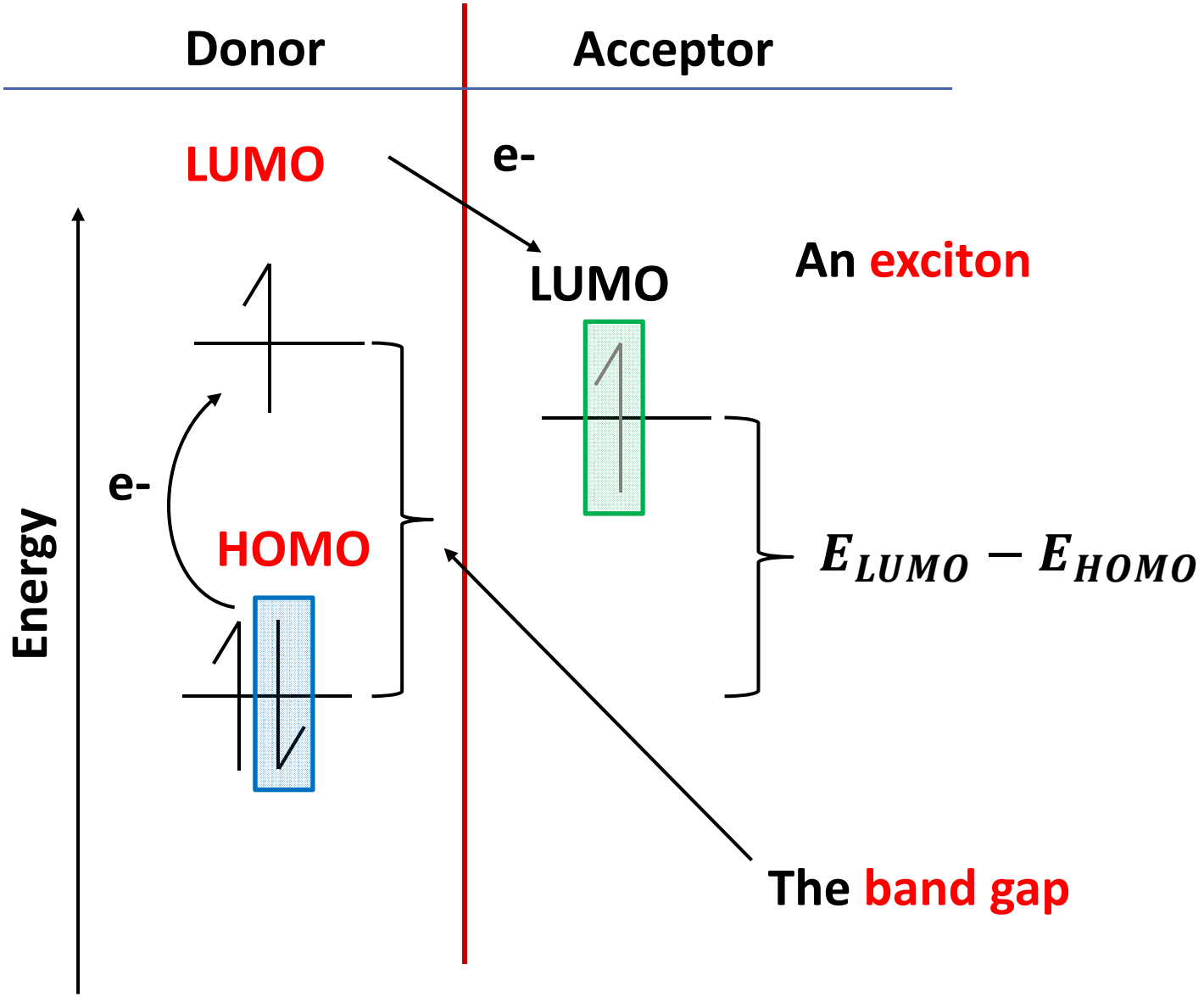

Hideki Shirakawa

HOW ORGANIC SOLAR CELLS WORK



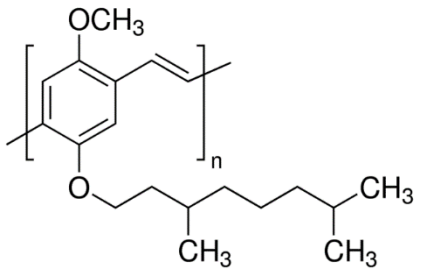


HOW ORGANIC SOLAR CELLS WORK

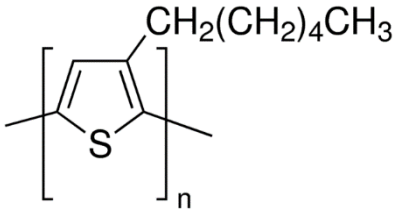


1. Exciton creation via photon absorption
2. Exciton diffusion to junction
3. Charge separation
4. Transport of charge to electrodes

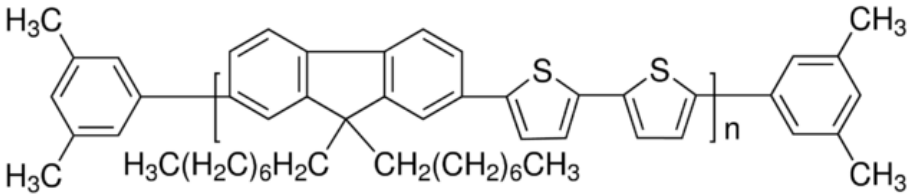
TYPICAL MATERIALS USED



MDMO-PPV

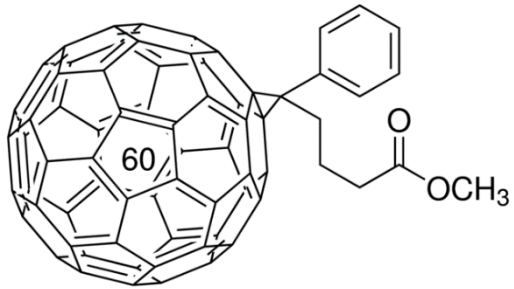


P3HT

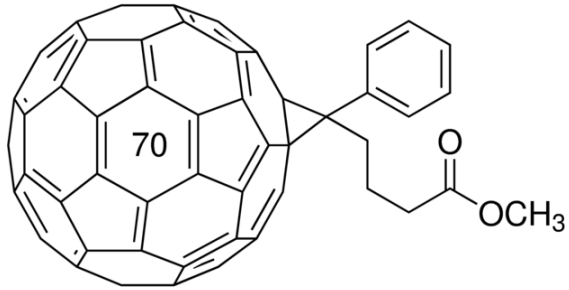


F8T2

Donors



PCBM - 60



PCBM - 70

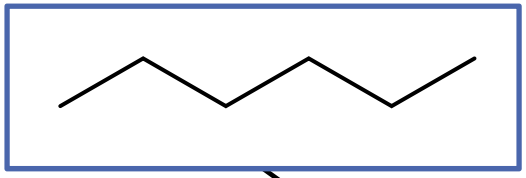
Acceptors

AN OPEN PROBLEM

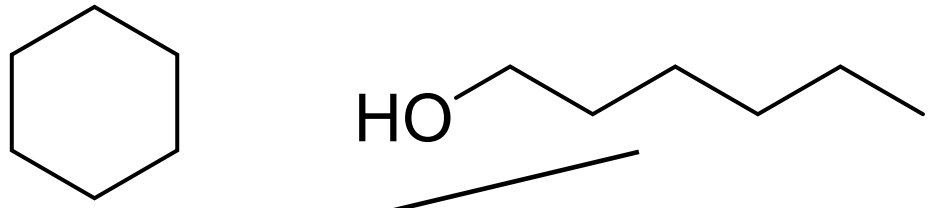
- **What is the composition and structure of the "best" performing organic solar cells?**
- **Performance metrics:**
  - **Efficiency**
  - **Stability**
    - ***Lifetime of silicon solar cells: 20 years***
  - **Production cost**
  - **Pay out time**
    - ***Silicon solar cells: 2.5 to 4 years***
  - **Ease of manufacture**
  - **Amount of CO_2 released during manufacture**
    - ***Silicon solar cells: 20 g per kWh produced***
  - **Amount of toxic material used during manufacture**
  - **Biodegradability and other environmental effects**

THE NECESSITY OF OPTIMIZATION

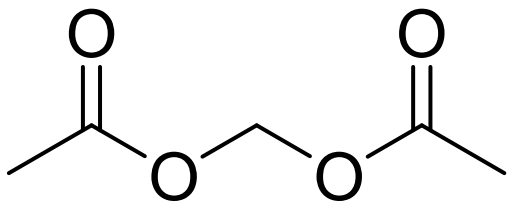
If we have some known-to-work compound:

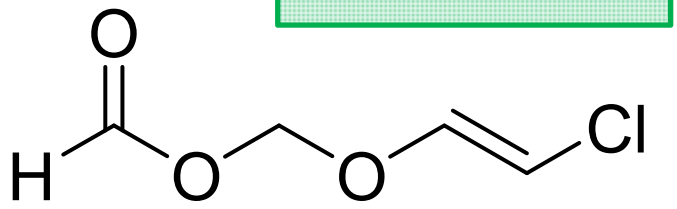


We might consider similar structures in designing a replacement:



However, this **trial-and-error approach fails to account for the molecular diversity of the complete chemical design space**. Using optimization, we can design:

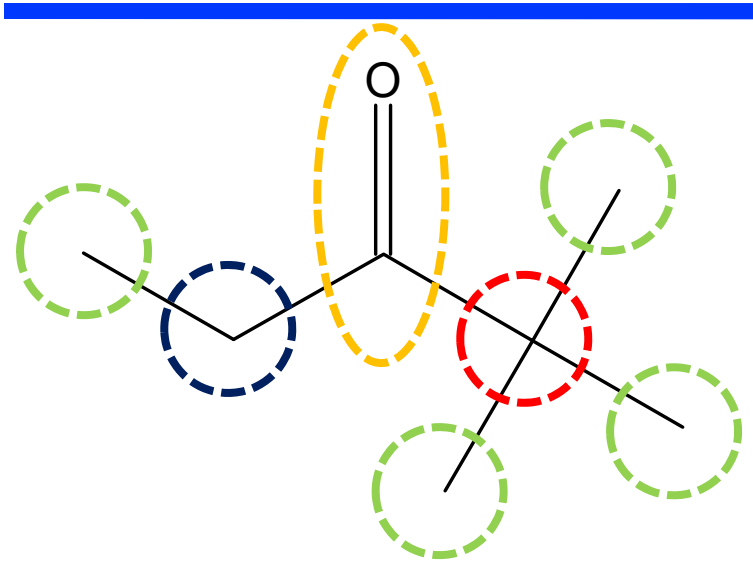




This example comes from a crystallization solvent case study (ibuprofen)

These vastly improved structures are not obvious from our known-to-work compound

GROUP CONTRIBUTION (GC) METHODS



$$n = [0, \ldots, 4, \ldots, 1, \ldots, 1, 0, 1, \ldots]$$

n	a vector of group occurrences
c	property-specific coefficients for each group

$$c = [\ldots, 12.53, \ldots, 0.54, \ldots, 5.12, 2.10, -1.55, \ldots]$$

$$P = f\left(\sum_i c_i n_i\right) = 4 \times 12.53 + 1 \times 0.54 + 1 \times 5.12 + 1 \times -1.55 = 54.23$$

if $f(x) = x$

MOLECULAR DESIGN

$$\begin{aligned} \max \quad & f(x,n) \\ \text{s.t.} \quad & g(x,n) \leq 0 \quad h(n) \leq 0 \\ & x \in \mathbb{R}^m \qquad n \in \mathbb{Z}^N \end{aligned}$$

n_i Frequency of group i

$h(n)$ Structural constraints on n

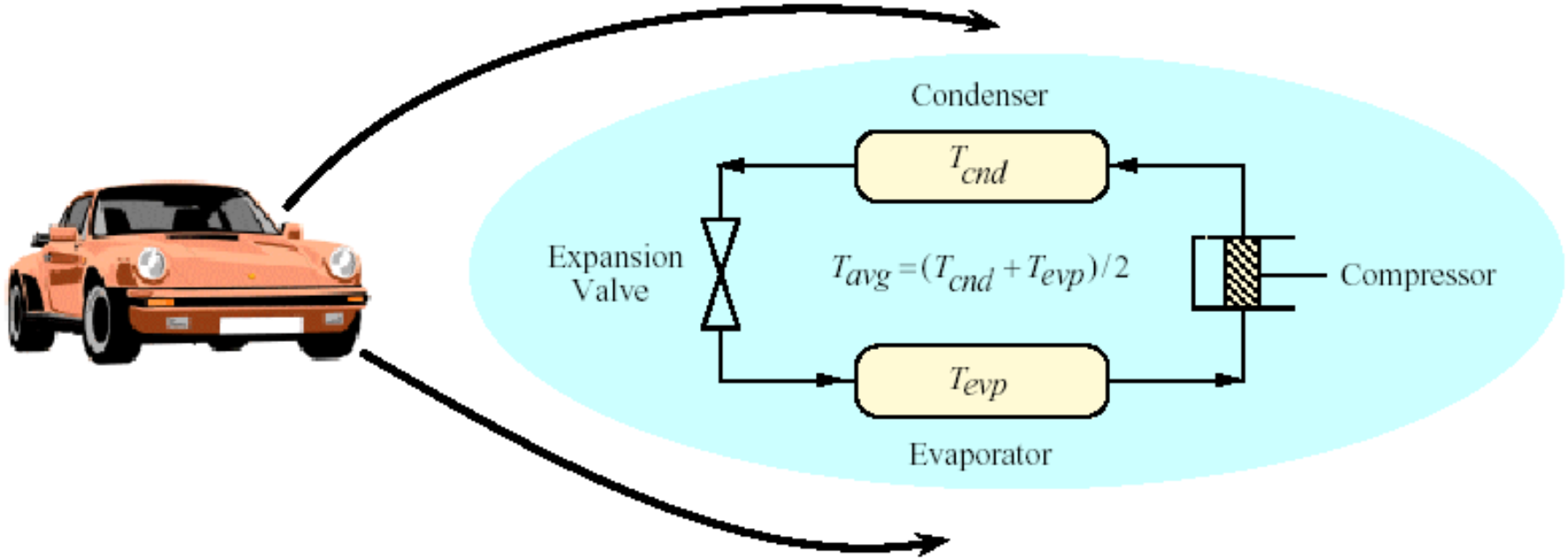
$g(x,n)$ Property estimation via GC and property bounds

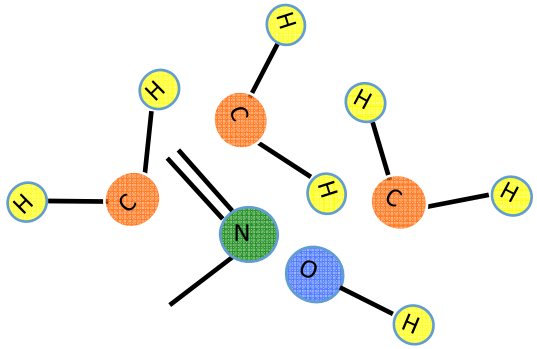
x Thermodynamic property

$f(x,n)$ Objective function

- **Odelle and Macchieto, Joback and Stephanopoulos, Gani and co-workers, Achenie and co-workers, Adjiman and co-workers, Sahinidis and co-workers**
- **The formulation is highly non-linear (properties; thermo)**

AUTOMOTIVE REFRIGERANT DESIGN

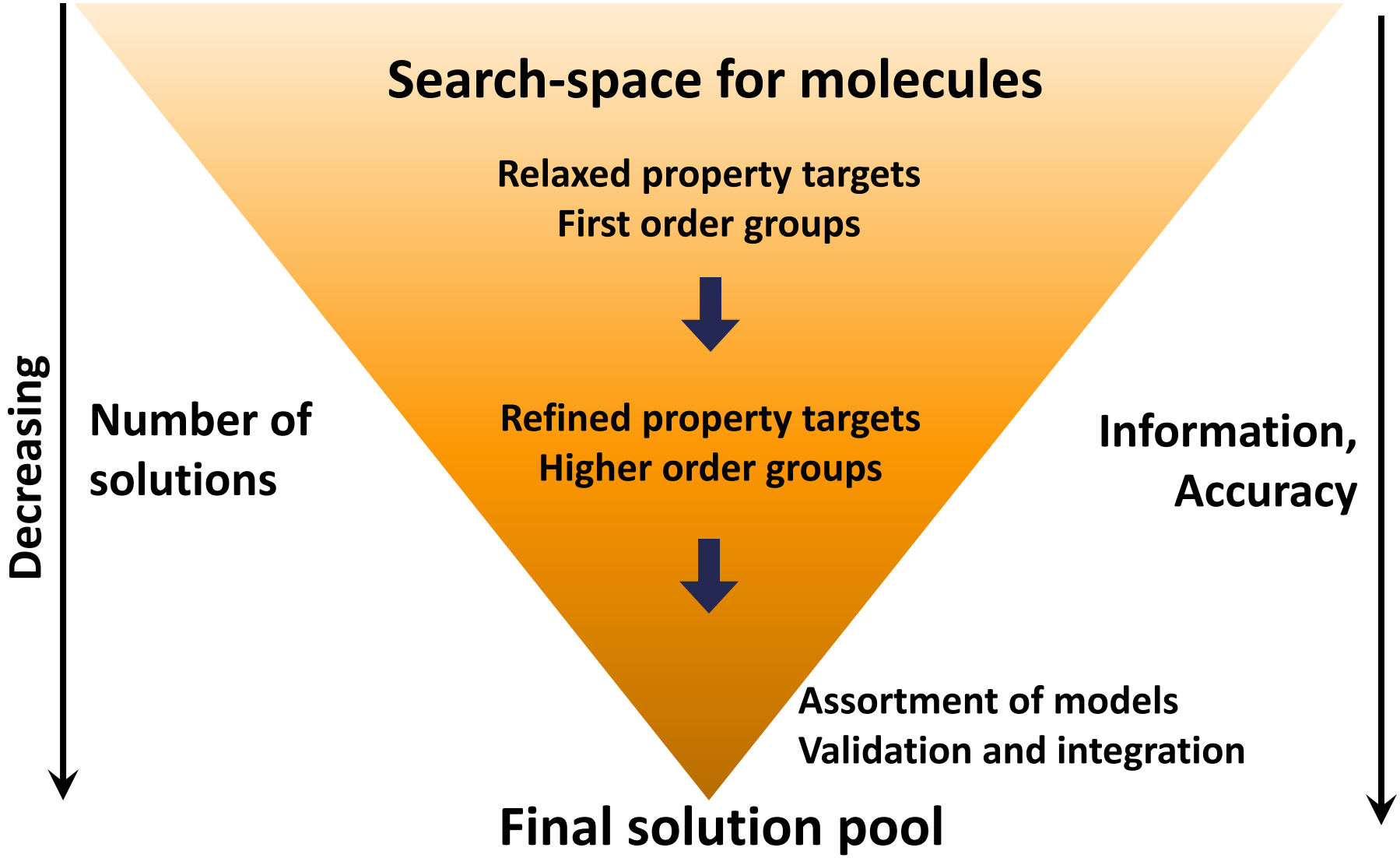




- **Higher enthalpy of vaporization (ΔH_{ve}) reduces the amount of refrigerant**
- **Lower liquid heat capacity (C_{pla}) reduces amount of vapor generated in expansion valve**
- **Maximize ΔH_{ve}/ C_{pla}**

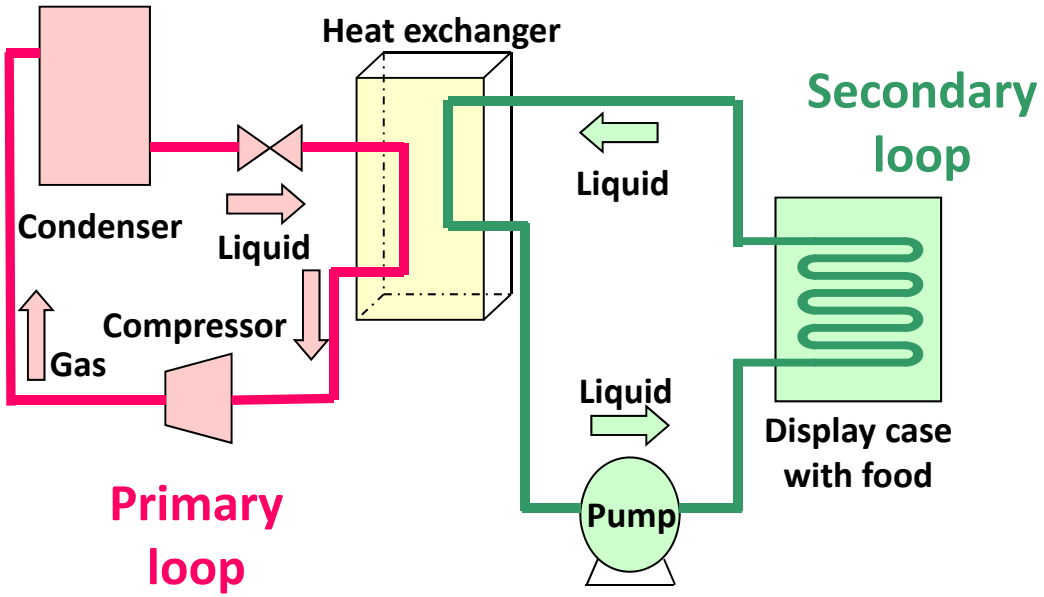
Sahinidis, Tawarmalani and Yu, 2003

AMODEO

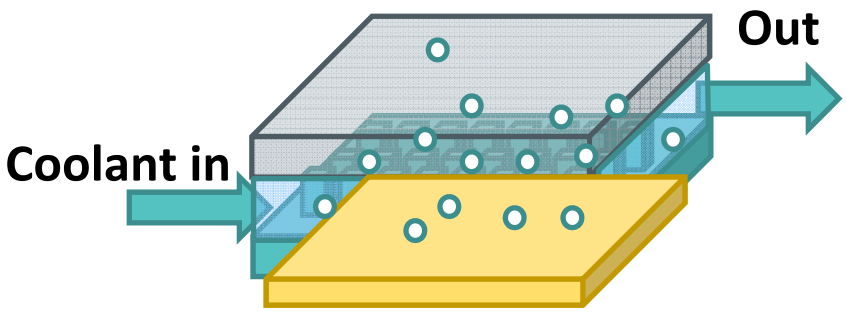


Samudra and Sahinidis, 2013

COOLANT DESIGN



Samudra and Sahinidis, 2013



Sun and Sahinidis, 2018+

MIXTURE MODELING

Project onto component property space

CAMxD problem (MINLP)

$$
\begin{aligned}
\min_{n,x} \quad & C(q) \\
\text{s.t.} \quad & q = g(x,p) \\
& p = f(n) \\
& h(x,p,q) \le 0 \\
& l(x,p,q) = 0 \\
& \sum_i x_i = 1 \\
& p^L \le p \le p^U \\
& q^L \le q \le q^U \\
& n \in \mathcal{S}
\end{aligned}
$$

Normal solution

$n^* = [\ldots, 1, \ldots, 3, \ldots, 1, \ldots]$

$x^* = [0.12, 0.58, 0.30]$

Too high-dimensional, non-linear, non-convex

Projection approach

Molecular structures

Much lower-dimensional; can easily design structures from properties

UNIFAC

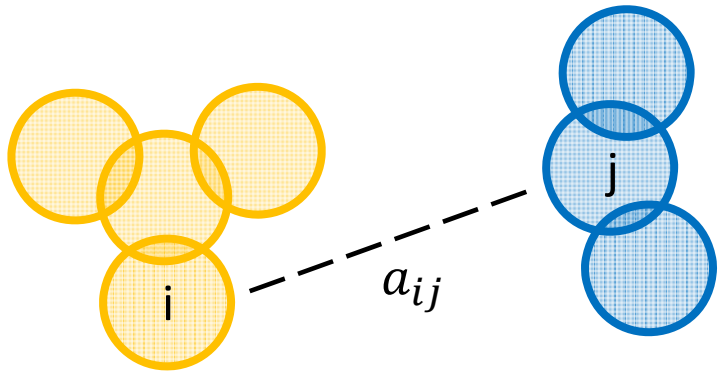
Activity coefficient estimation

$$\gamma = \mathrm{UNIFAC}(x, T, q, r)$$

Group contribution

$$r = \sum_i c_i^r n_i \qquad q = \sum_i c_i^q n_i$$

Binary interaction parameter table

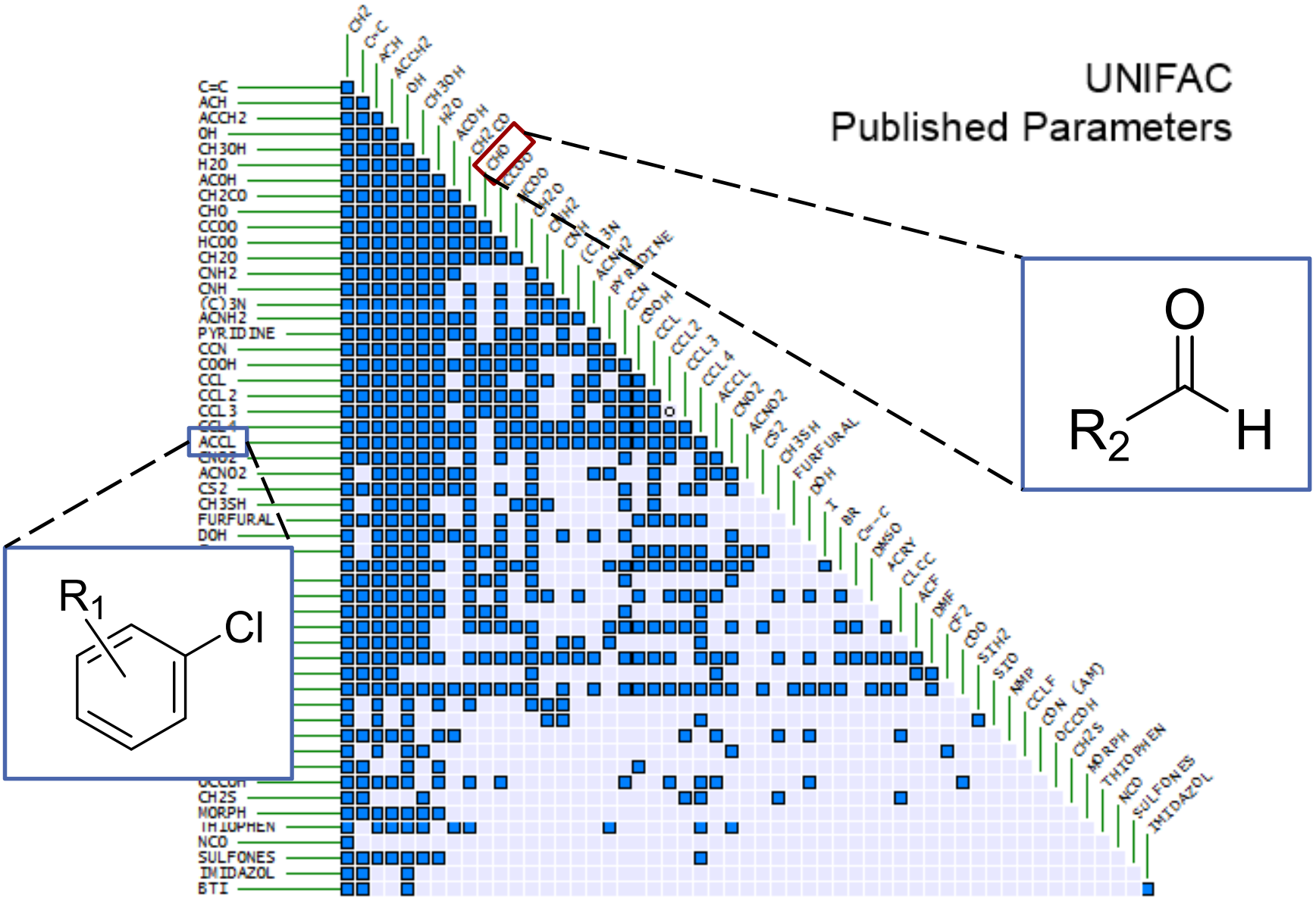


What if we do not have interaction parameters for every pair of groups in solution?

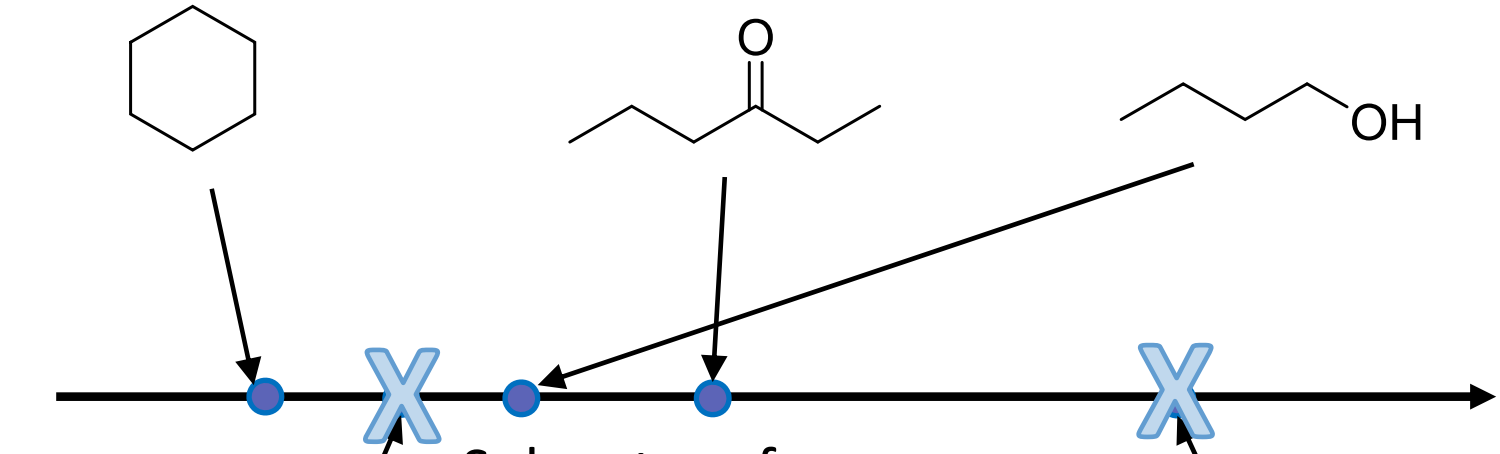
We cannot use UNIFAC.

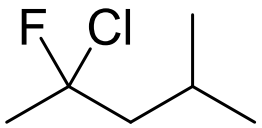
Consequence: Chemical design space is inherently limited

UNIFAC INTERACTIONS

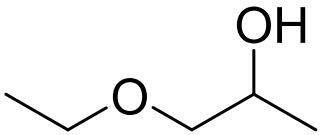


WHEN CAN THIS MAKE A DIFFERENCE?



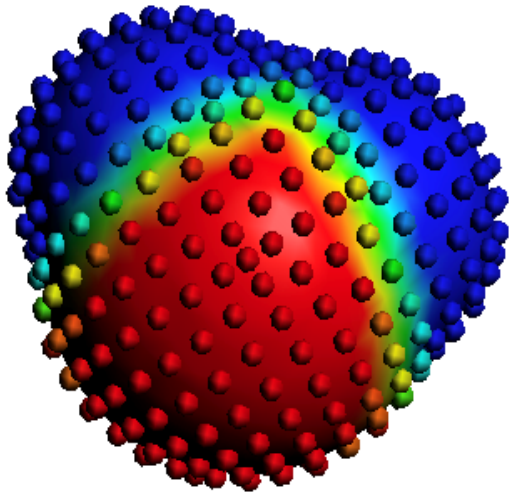


Activity coefficients for these two molecules cannot be predicted for this problem with UNIFAC

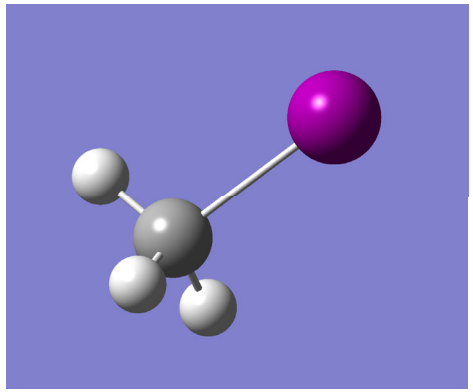


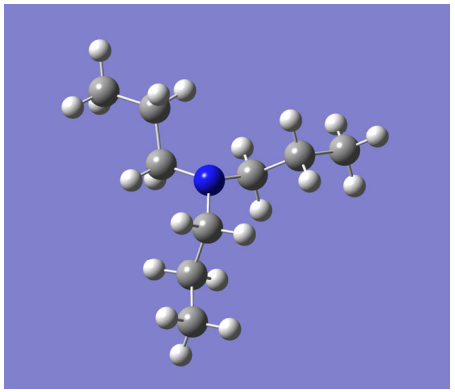
This comes from an extraction solvents design problem (acetic acid)

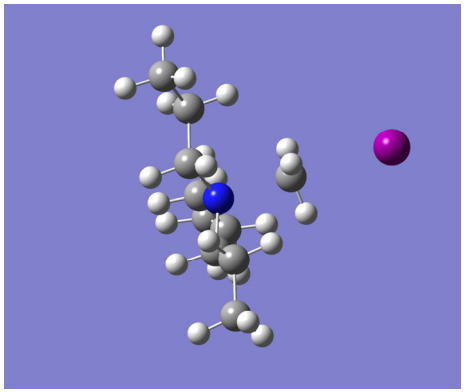
COSMO-based CAMD

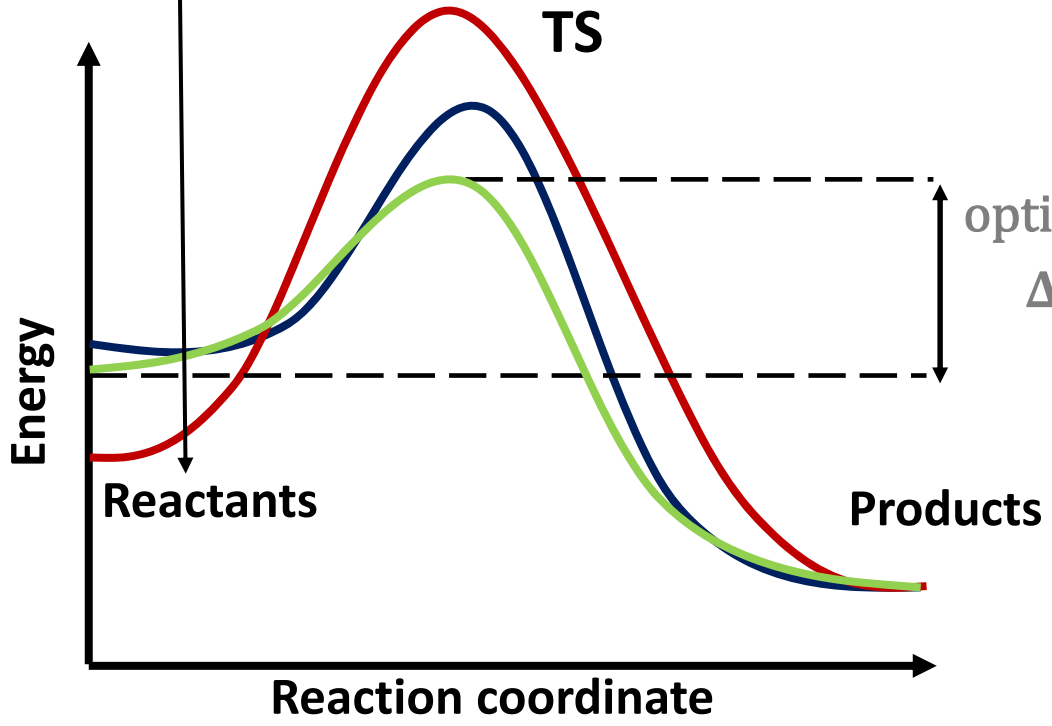


INTEGRATING QM INTO CAMD



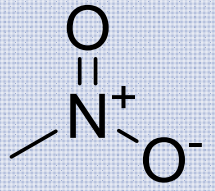






$\Delta G^{\ddagger}$ in different solvents calculated with **COSMO-RS**

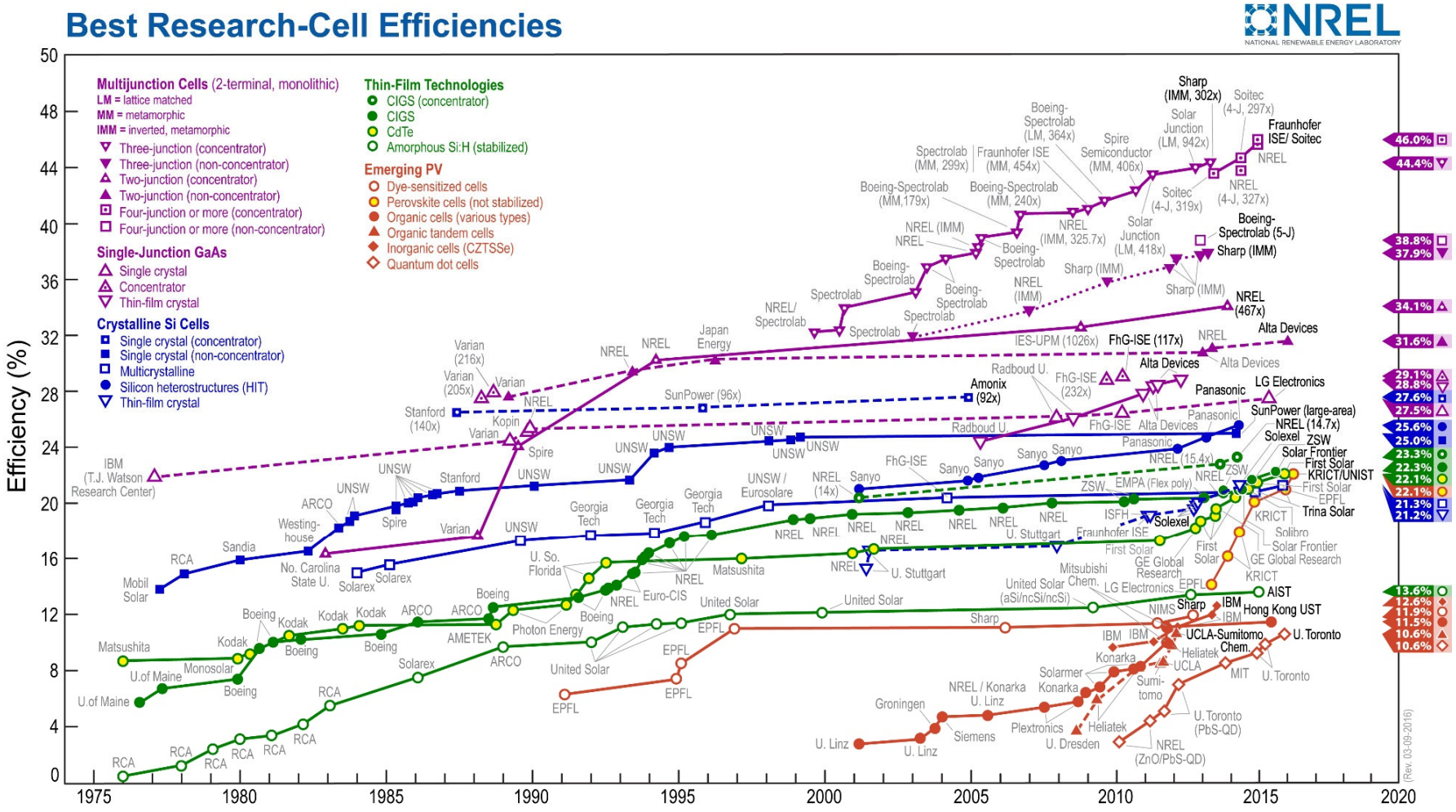
Optimal reaction solvent:



ANOTHER OPEN PROBLEM

- **Develop an accurate expression of efficiency of the solar cell in terms of**
  - **Molecular structure**
  - **Material morphology**
  - **Cell design**

BEST RESEARCH CELL EFFICIENCIES



WHY ORGANIC PHOTOVOLTAICS?

Advantages:

- Easy and inexpensive to produce
- Do not require Cd or other heavy metals
- Recyclable
- Lightweight and flexible with broader applications

Challenges:

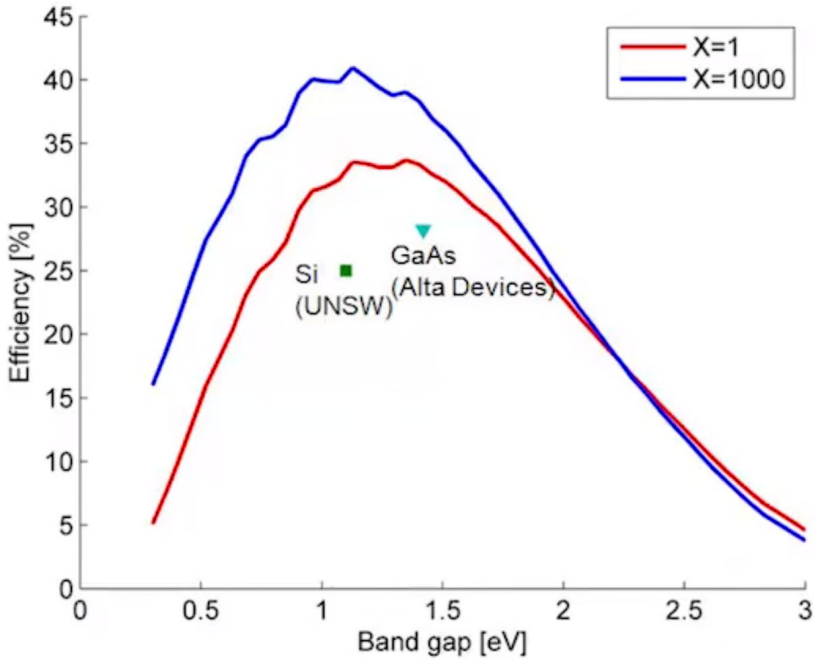
- Efficiency is much lower (10%) compared to silicon solar cells (~25-30%)
- Economic feasibility threshold is normally thought to be 15% efficiency
- Degradation and weathering lead to average lifetimes of 2 years

THERMODYNAMIC LIMITS

- Solar cell at $T_C \cong 300$ K
- Sun at $T_H \cong 6000$ K
- Carnot limit: $1 - 300/6000 = 95\%$

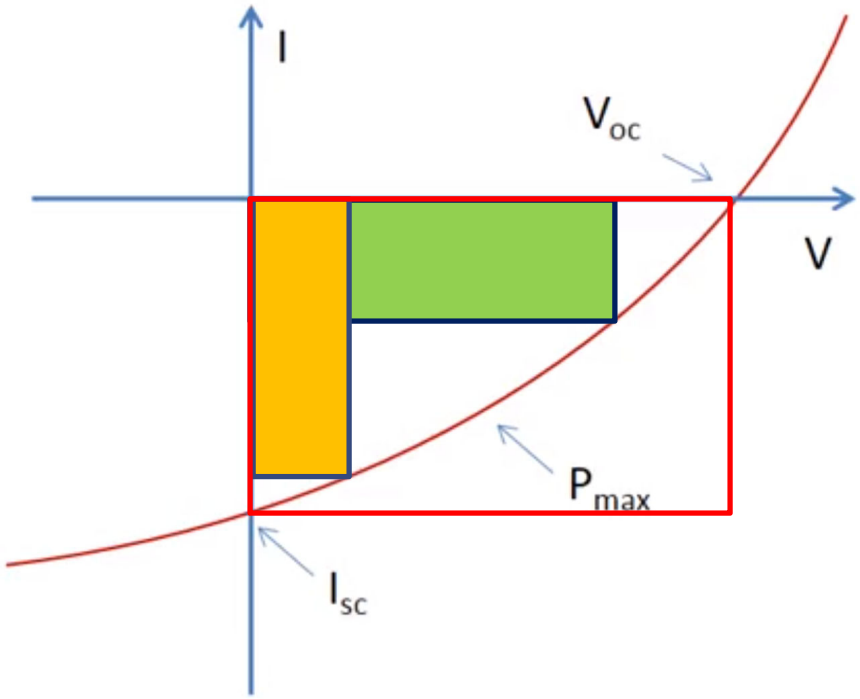
- Planck's law: 85.4%

- Shockley-Queisser (1961)



MODELING EFFICIENCY

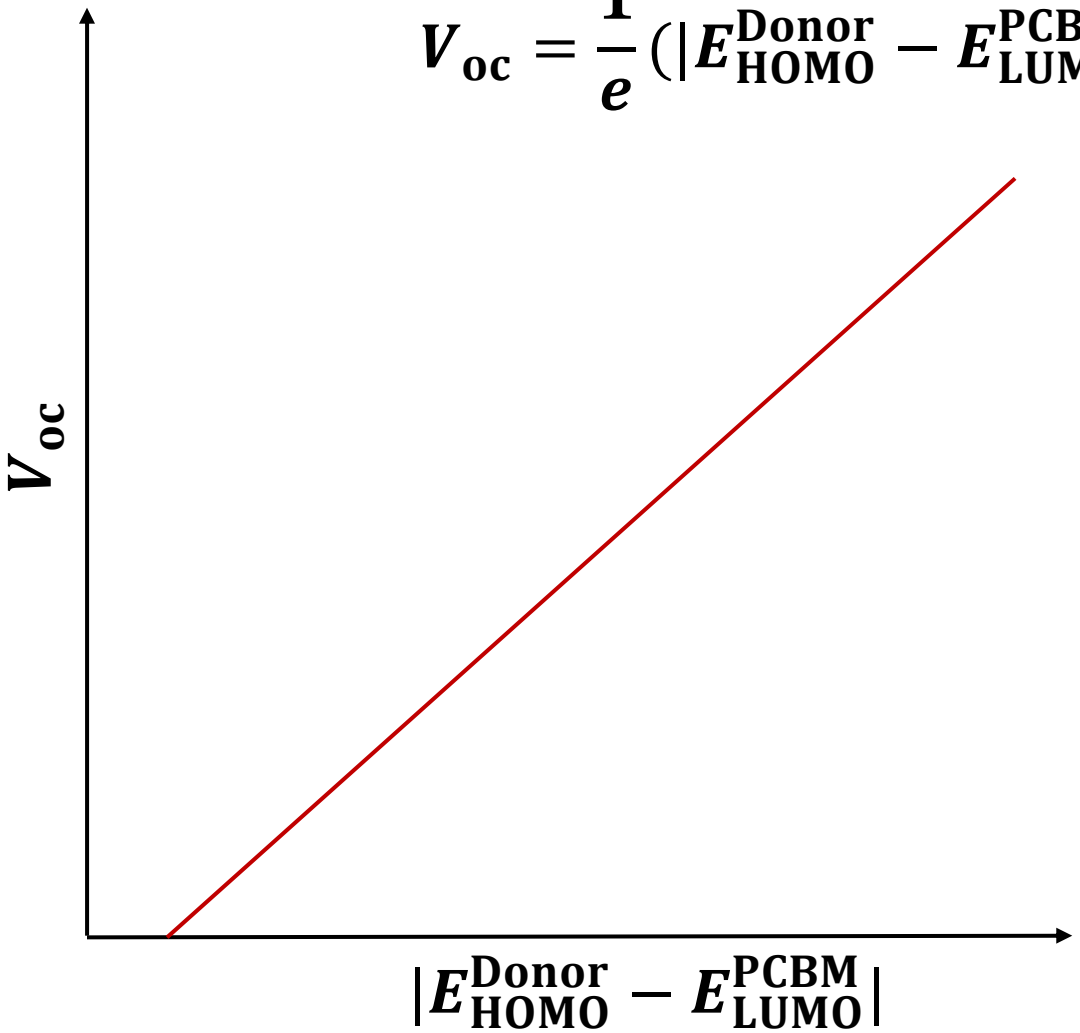
- $\eta = \frac{P_{max}}{P_{in}}$
- $P = I \times V$
- $\eta = \frac{FF \times J_{sc} \times V_{oc}}{P_{in}}$

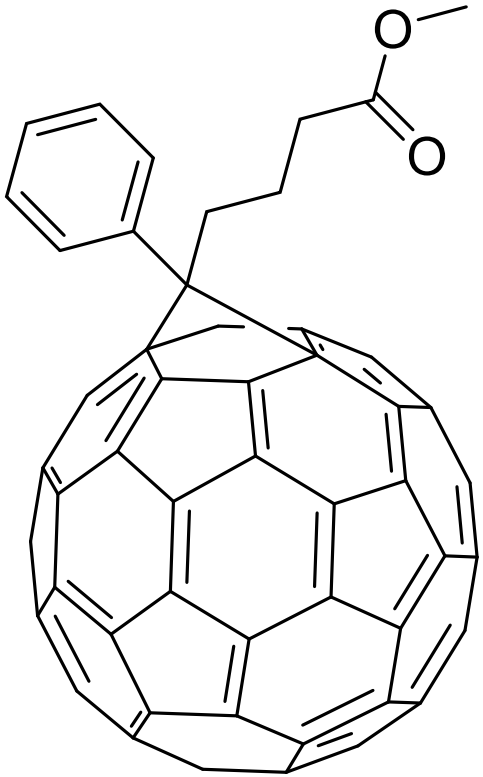


- η = efficiency (PCE)
- I = current
- V = voltage
- P = power
- SC: short circuit
- OC: open circult
- FF: fill factor

SCHARBER MODEL

$$V_{oc} = \frac{1}{e}(|E_{HOMO}^{Donor} - E_{LUMO}^{PCBM}|) - 0.3V$$





- **PCBM as the acceptor**
- **Data from 26 devices**

Scharber et al., 2006 (Konarka Austria)

HARVARD CLEAN ENERGY PROJECT

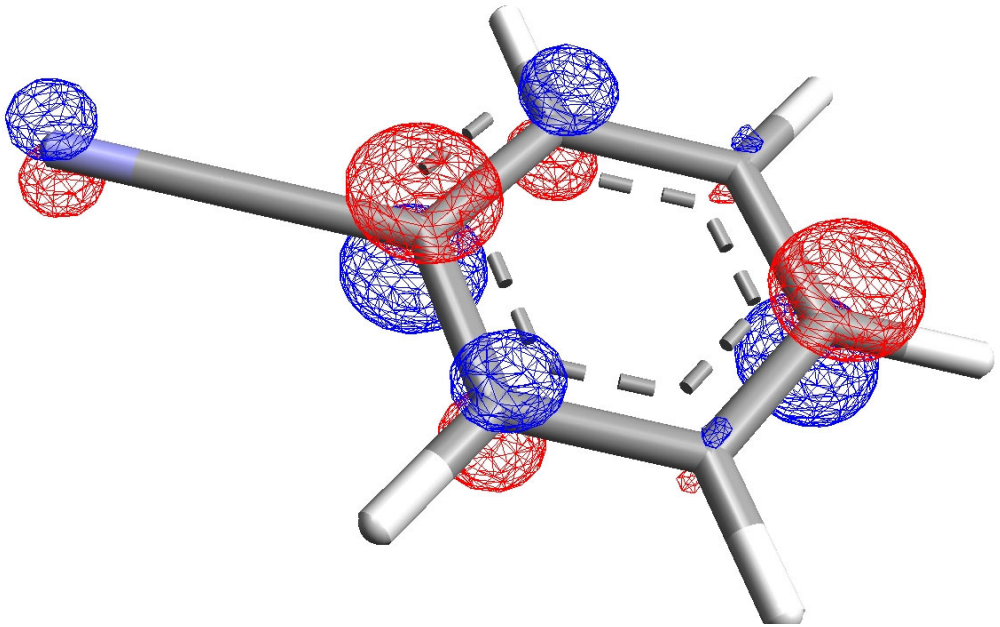
Basic procedure

1. **Use SMILES to connect building blocks in every possible way**

2. **Perform DFT calculations to determine HOMO/LUMO levels of donor materials (using PCBM as the acceptor)**

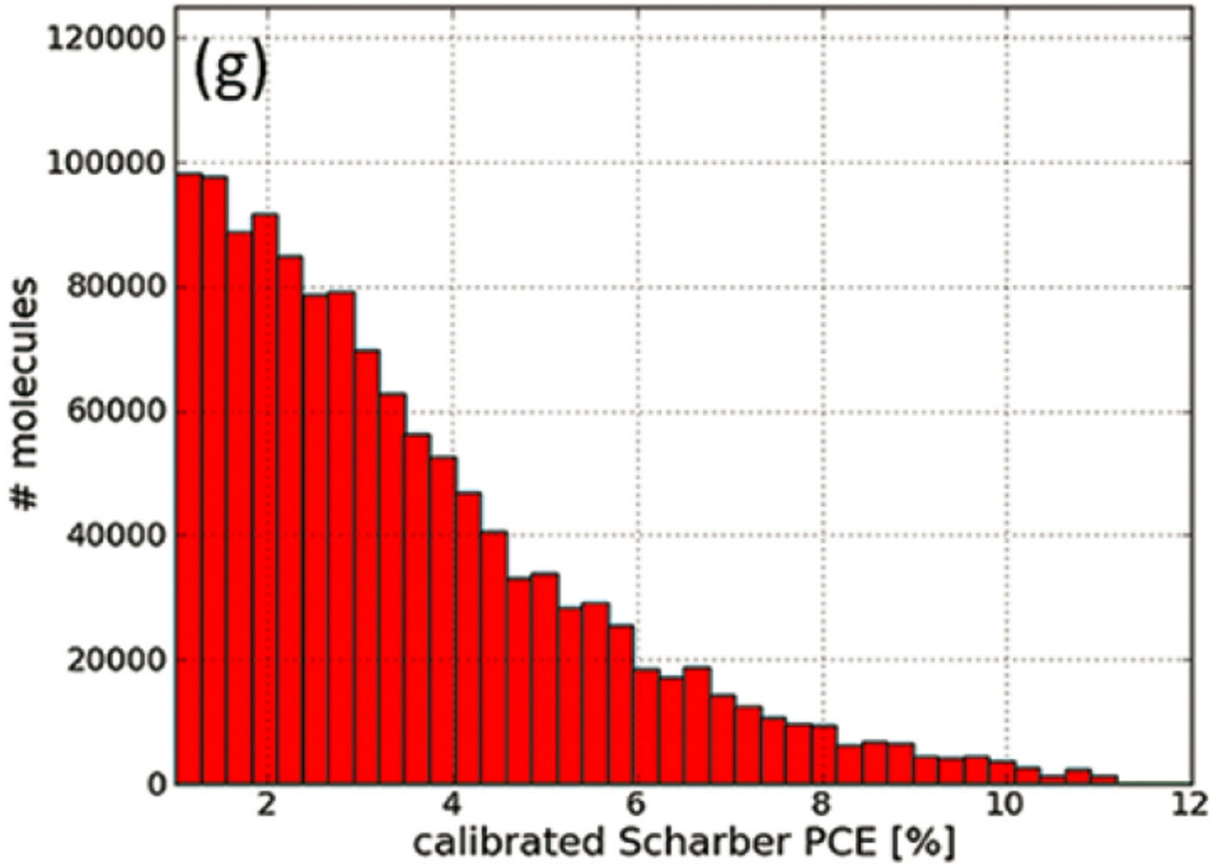
3. **Calculate efficiency with the Scharber model**

Most difficult and time-consuming step



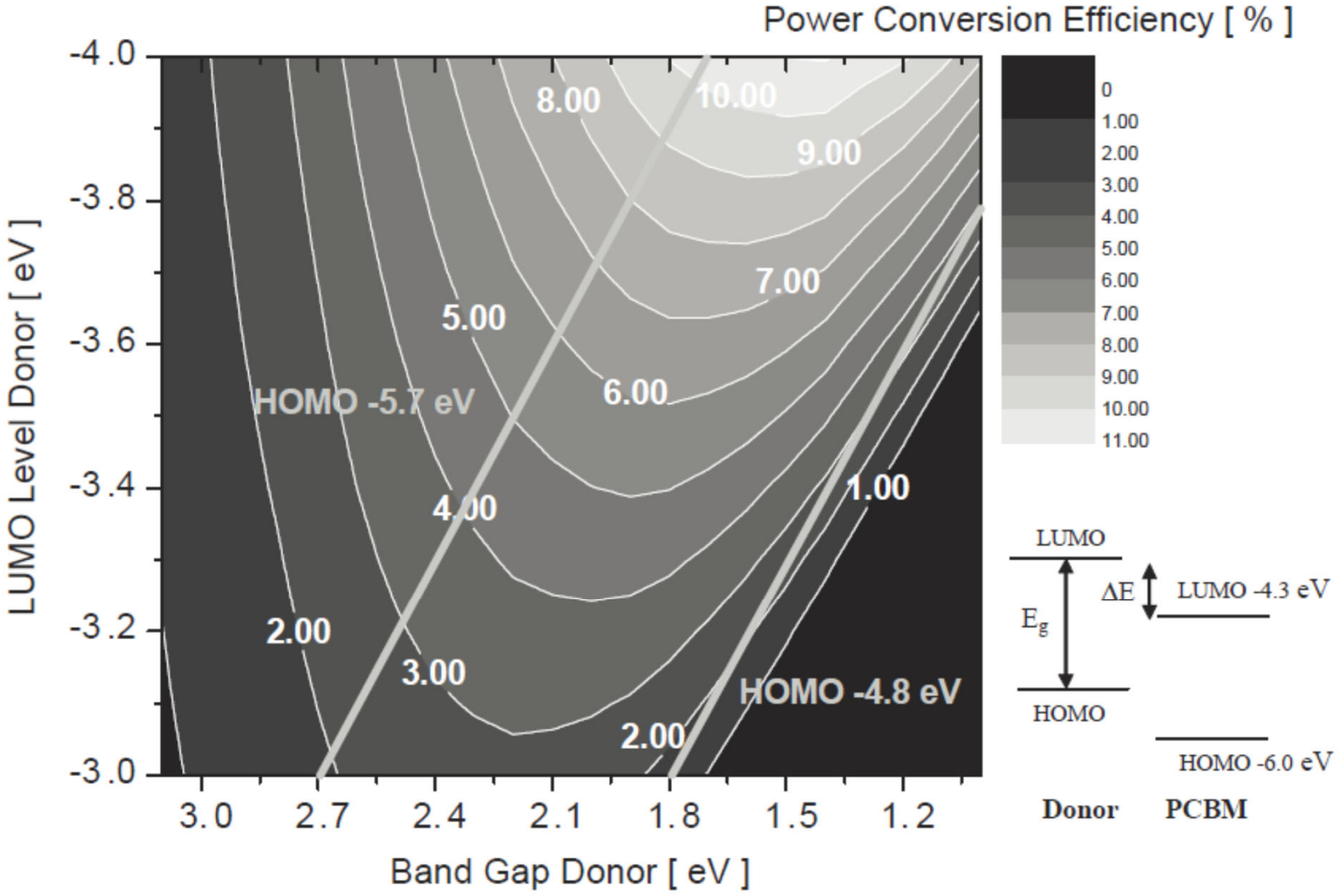
Aspuru-Guzik et al., 2011

RESULTS FROM 2.6 MILLION DESIGNS



Aspuru-Guzik et al., 2011

MAJOR LIMITATION



Scharber et al., 2006

ISSUES WITH THE SCHARBER MODEL

- **Fixed *FF* at 0.65**
- **Equation to calculate V_{oc} is based on two parameters only**
- **Model built from 26 points only**
- **Developed specifically for acceptor PC61BM**
- **Theoretical limit on efficiency of around 11%**
  - **Experimental cells have shown to have higher efficiencies**

MODELS WITH GC METHODS

SMARTS up to size 5

Property	Number of groups	Number of interaction terms	Number of structural features	R^2
V_{oc}	26	1	5	0.72
J_{SC}	26	3	5	0.70
FF	23	5	5	0.67
η	24	5	5	0.77

SMARTS up to size 10

Property	Number of groups	Number of interaction terms	Number of structural features	R^2
V_{oc}	81	4	5	0.75
J_{SC}	81	5	5	0.72
FF	81	5	4	0.68
η	81	5	5	0.78

Using 349 measurements

MODELS USING QUANTUM DESCRIPTORS

- **Linear fitting with descriptors from DFT calculations**
  - Molecular mass, electronegativity, electron density, van der Waals radii, polarizability, hydrogen bond count, steric hindrance

Property	No. of descriptors	R^2
V_{oc}	20	0.9491
J_{SC}	16	0.8917
FF	20	0.6402
η	15	0.8460

OPEN PROBLEMS

1. **Accurate prediction of efficiency based on molecular structure, morphology, processing**
   - Build computational models from first principles
   - Use ALAMO to build simple, accurate nonlinear models
     - *Lack of data*
       - Harvard group (Aspuru-Guzik et al.)
       - Denmark group (Krebs et al.)
   - Synthesize and measure
2. **Modeling of stability, environmental impact, production costs/constraints**
3. **Design of optimal organic photovoltaic cell**
4. **Energy storage**