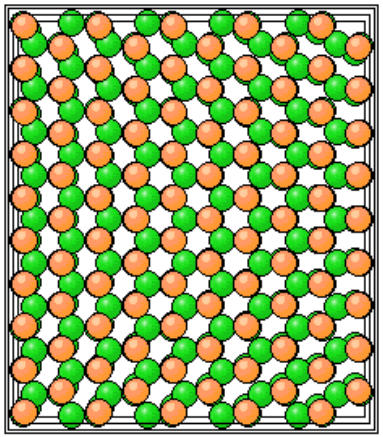
Derivative-Free Optimization Methods for a Surface Structure Determination Problem

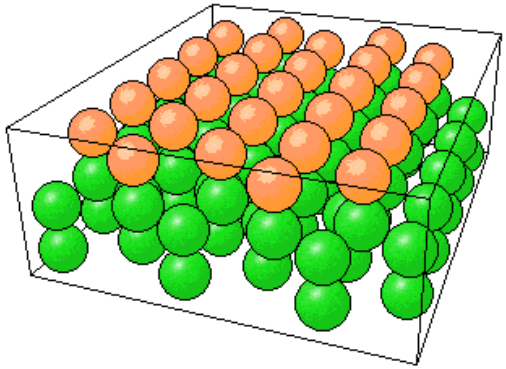
Juan C. Meza
University of California, Merced

Applied Math Seminar
October 7, 2016 UCM

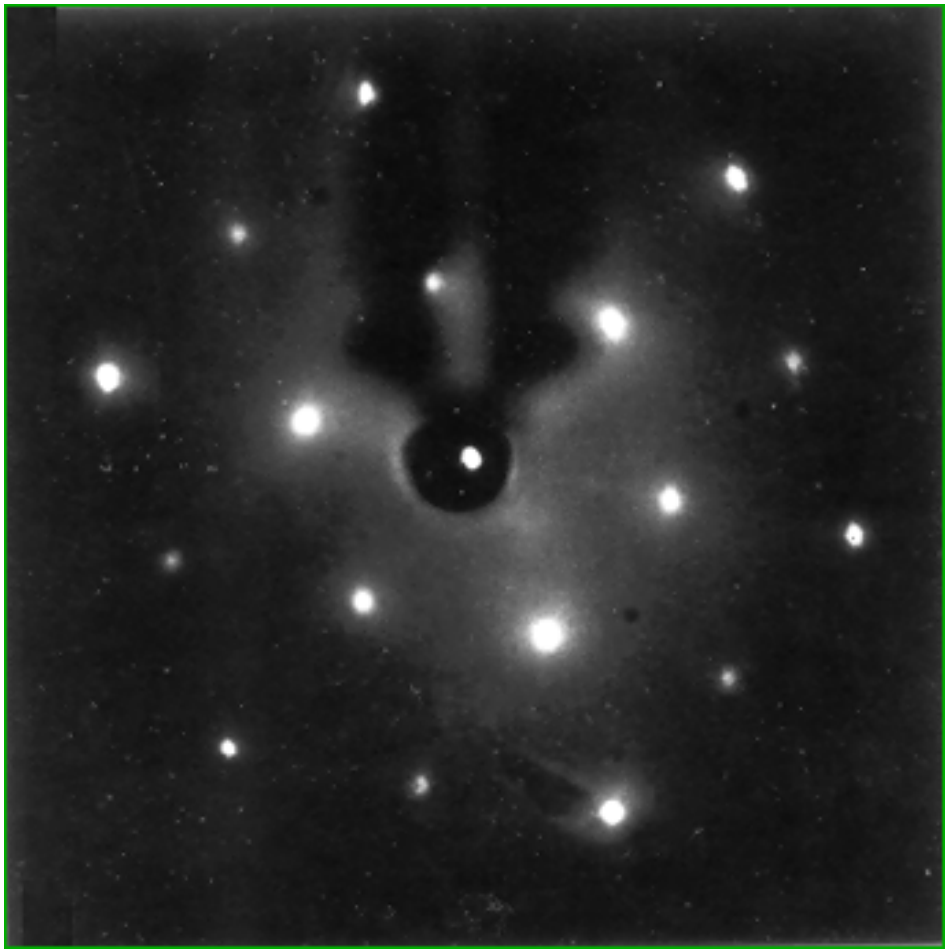
Surface structure determination

- Surface structure determines many important properties of materials
- Possibility of electron diffraction first proposed by deBroglie (1924)
- Experimental data could not be quantitatively described by kinematic theory
- Over 40 years before electron diffraction became a tool in surface structure determination
- Necessitated the development of a theory of multiple scattering (late 1960s)



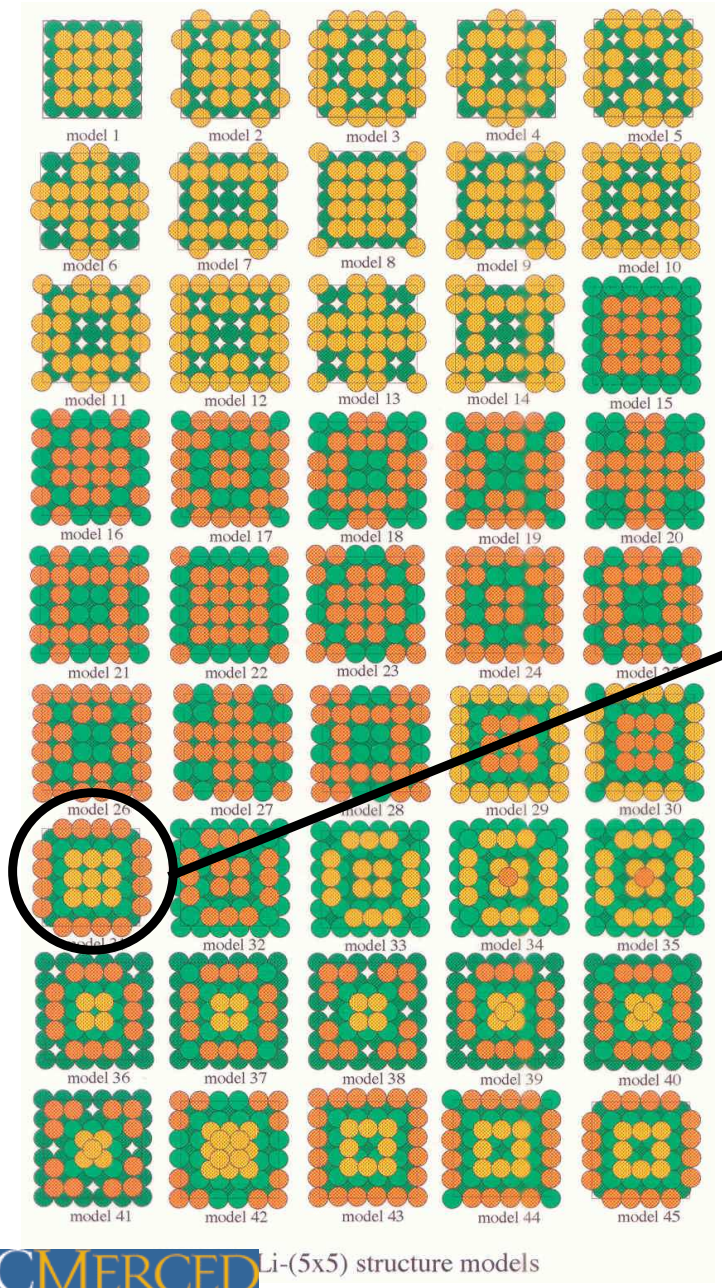


Low-energy electron diffraction (LEED)

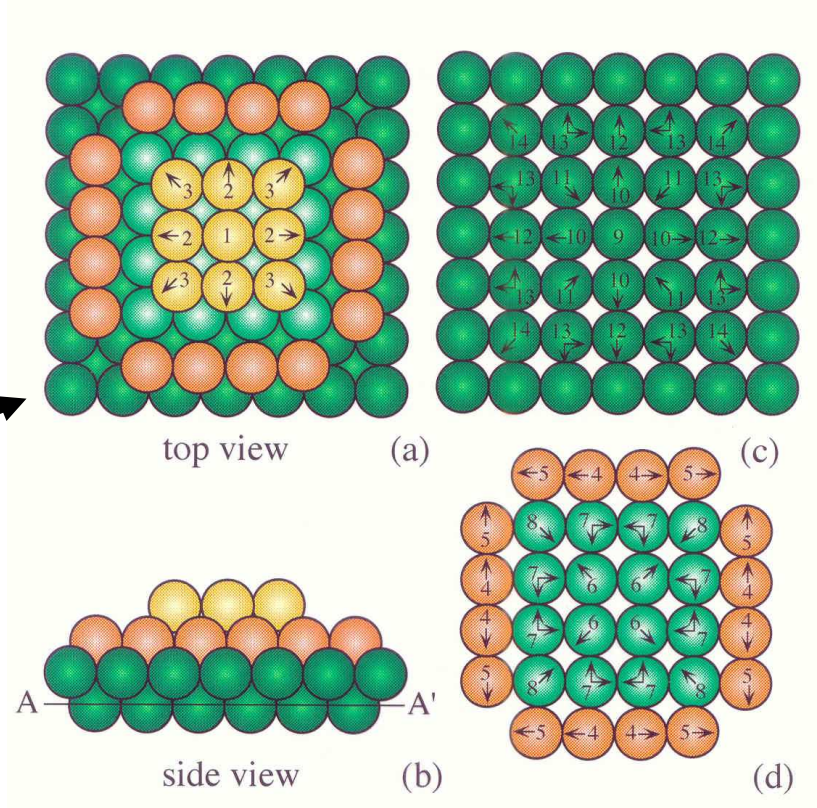


- Goal is to determine surface structure through low energy electron diffraction (LEED).
- Need to determine the coordinates and chemical identity of each atom.
- Ex: Li atoms on a Ni surface
- Low-energy electrons have high surface sensitive, requiring a well-ordered surface

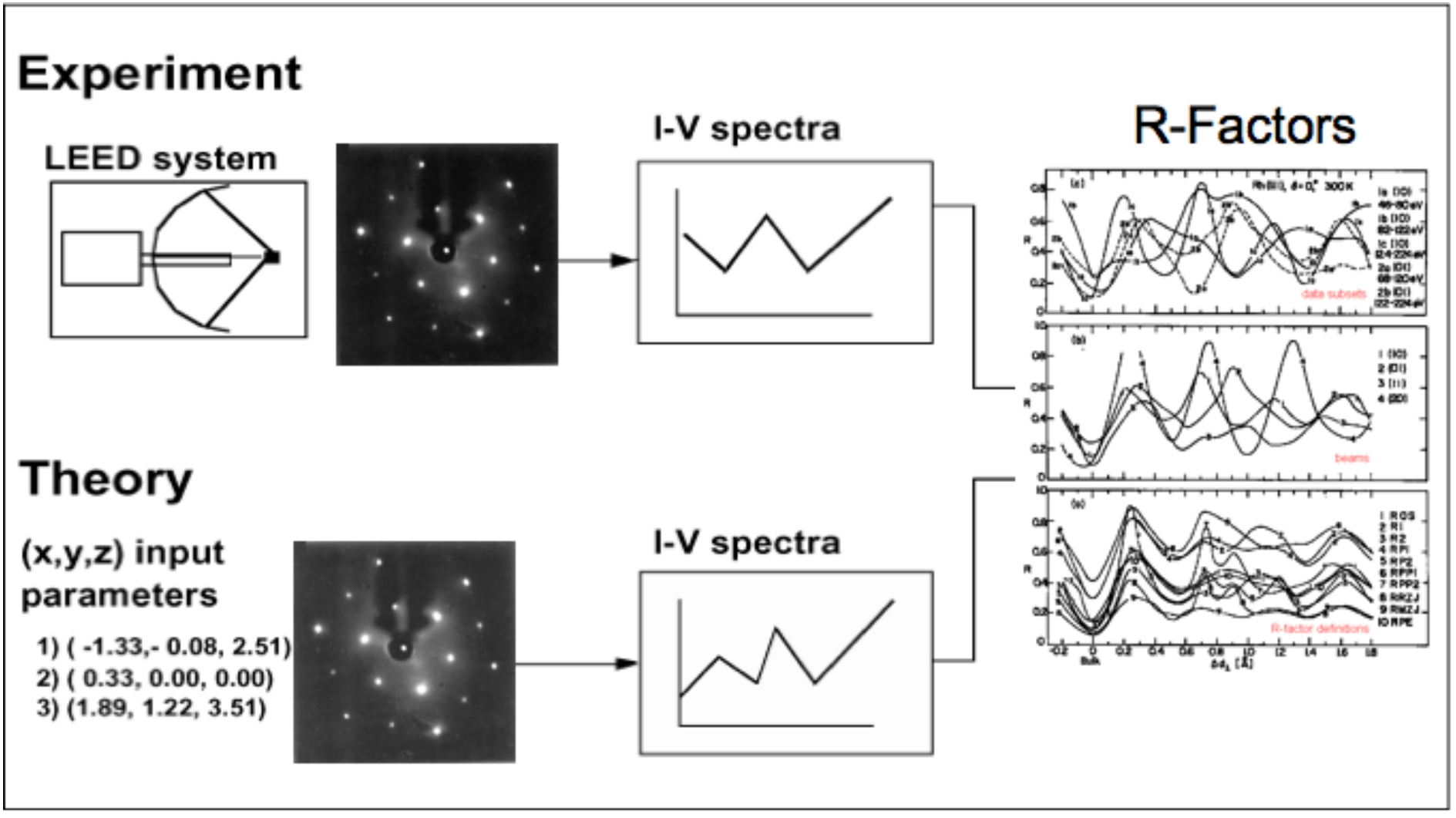
What is the correct atomic configuration?



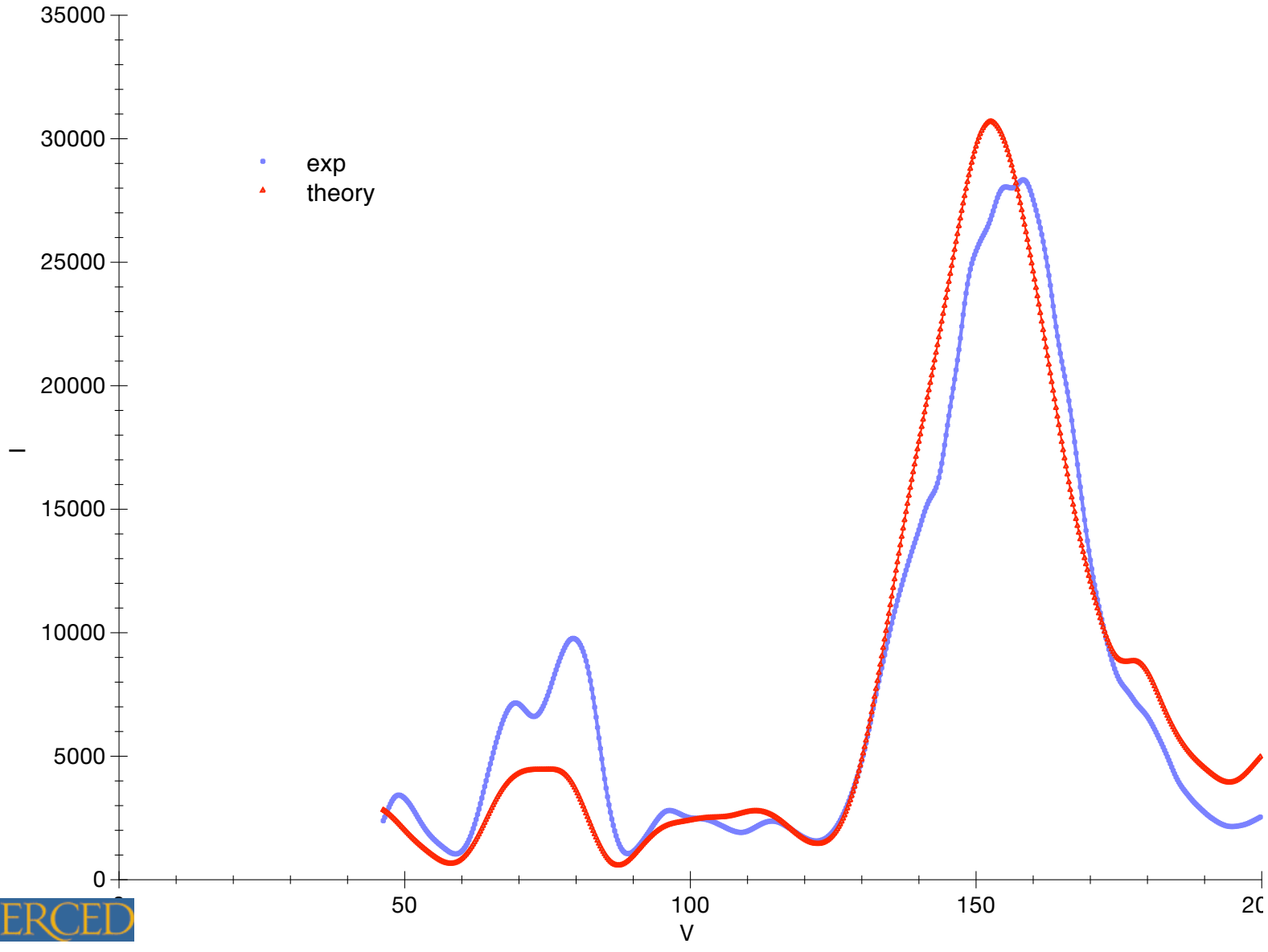
Li-(5x5) structure models



Experimental setup



Example IV curve



Pendry Reliability-factor (1980)

- IV curves consist for the main part of a series of Lorentzian peaks (I is Intensity):

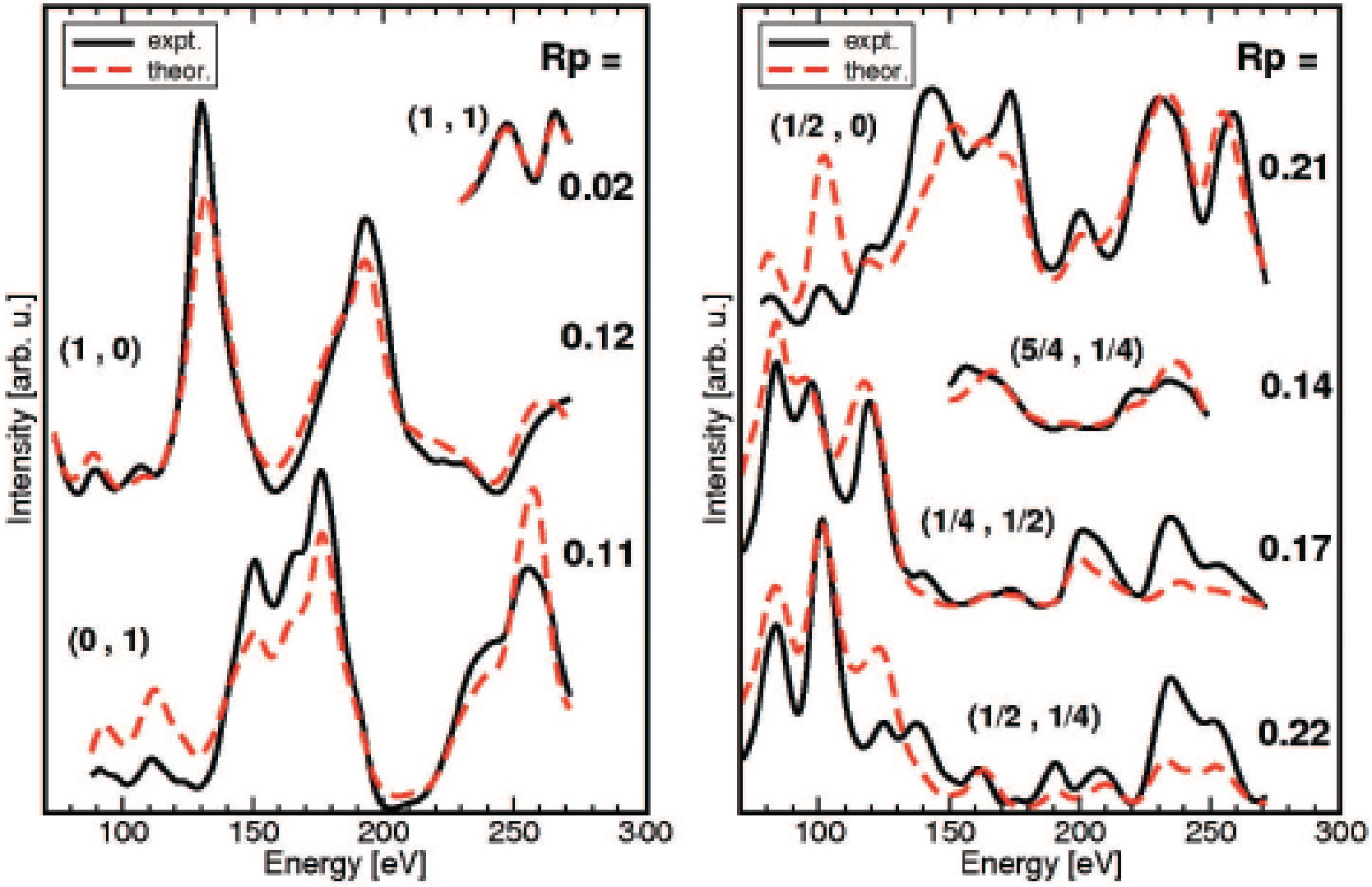
$$I \approx \sum \frac{a_j}{(E - E_i)^2 + V_{oi}^2}$$

- Pendry R-factor emphasizes positions of the peaks rather than the heights of the intensities
- Ideal agreement corresponds to $R = 0$; uncorrelated spectra yields $R = 1$.

$$\begin{aligned} L(E) &= I'/I \\ L &\approx \sum_i \frac{-2(E - E_i)}{(E - E_i)^2 + V_{oi}^2} \\ Y(E) &= L/[1 + (LV_{oi})^2] \\ R &= \frac{\sum_{i,g}(Y_{th} - Y_{exp})^2}{\sum_{i,g}(Y_{th}^2 + Y_{exp}^2)} \end{aligned}$$

where g denotes one beamline and E_i is the energy.

R factors for various IV curves



Held, G., Low-energy electron diffraction crystallography of surfaces and interfaces. Bunsen-Magazin 12 (12), 2010. pp. 124–131.

General Optimization Formulation

$$\begin{aligned} min \qquad & f(x), \qquad x \in R^n \\ & h(x) = 0 \\ & g(x) \geq 0 \end{aligned}$$

For our problem:

- $f(x) =$ R-factor, x are coordinates of the atoms
- Several ways of computing the R-factor, we will use the Pendry R-factor.
- Sometimes referred to as parameter estimation or inverse problem

Some standard assumptions

- Theoretical
  - Objective function is smooth
  - Derivatives are also "nice"
  - Constraints are linearly independent and smooth
- Practical
  - First and (sometimes) second derivatives are available
  - Objective and constraint functions are cheap to evaluate
  - Objective function has infinite (machine) precision

Let's check our assumptions

- Function may be discontinuous
- Invalid (unphysical) structures lead to function being undefined in certain regions and returning special values
- No derivatives available; Hessians even harder to compute
- Objective and constraint function evaluations take up a substantial total of run time
- Function is a result of a computer simulation fitted to experimental data, i.e. less than infinite precision

And just for fun let's also add

- Combination of continuous and categorical variables
  - Atomic coordinates: x, y, z
  - Chemical identity: Ni, Li

Previous Work

- Early attempts used Hooke-Jeeves, nonlinear least squares, genetic algorithms[1−2]
- Effective, but expensive
  - Several 100,000s of function calls typically needed
  - Each function call can take several minutes on a workstation class computer
- We have also used pattern search methods – better, but still expensive

1. Global Optimization in LEED Structure Determination Using Genetic Algorithms, R. Dll and M.A. Van Hove, Surf. Sci. 355, L393-8 (1996).
2. G. S. Stone, MS dissertation, Computer Science Dept., San Francisco State University, 1998.

What is a Pattern Search Method?

- Particular instance of Direct Search method - methods that "in their heart" do not use gradient information, e.g. Nelder-Mead simplex (M. Wright, 1996).
- Can also be classified as a Generating Set Search method
- Main operation is function comparisons
- Useful whenever the derivative is not available or too expensive to compute, for example so called simulation-based optimization problems
- Unlike other DFO methods, these methods are strictly monotonic (vs. GA or SA)

General observations

- Use multiple search directions in such a way as to ensure at least one is a descent direction
- Makes the methods ideal for parallel computation
- Can use simple or sufficient decrease
- Never uses gradient in practice, but theory does require gradient be Lipschitz continuous or continuously differentiable

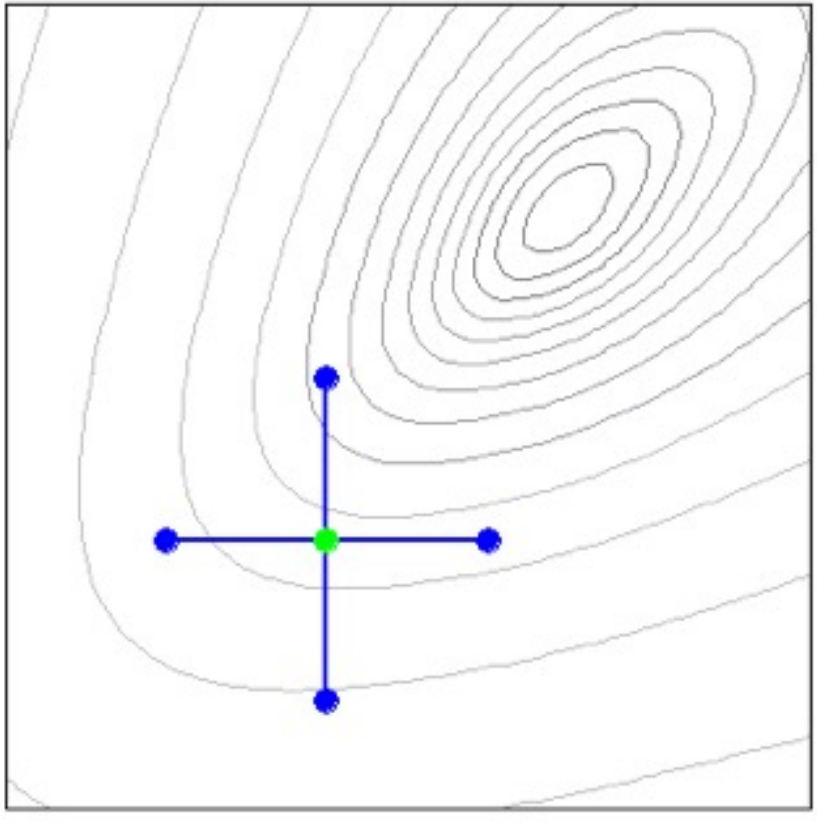
General Algorithm

Initialization: Given values Δ_0, x_0, M_0, P_0,

1 For $k = 0, \ldots,$ maxit do
2 Search: Evaluate f on a finite subset of trial points on a mesh M_k.
3 Poll: Evaluate f on the frame P_k.
4 If (successful)
5 $x_{k+1} = x_k + \Delta_k d_k$
6 Update Δ_k
7 End

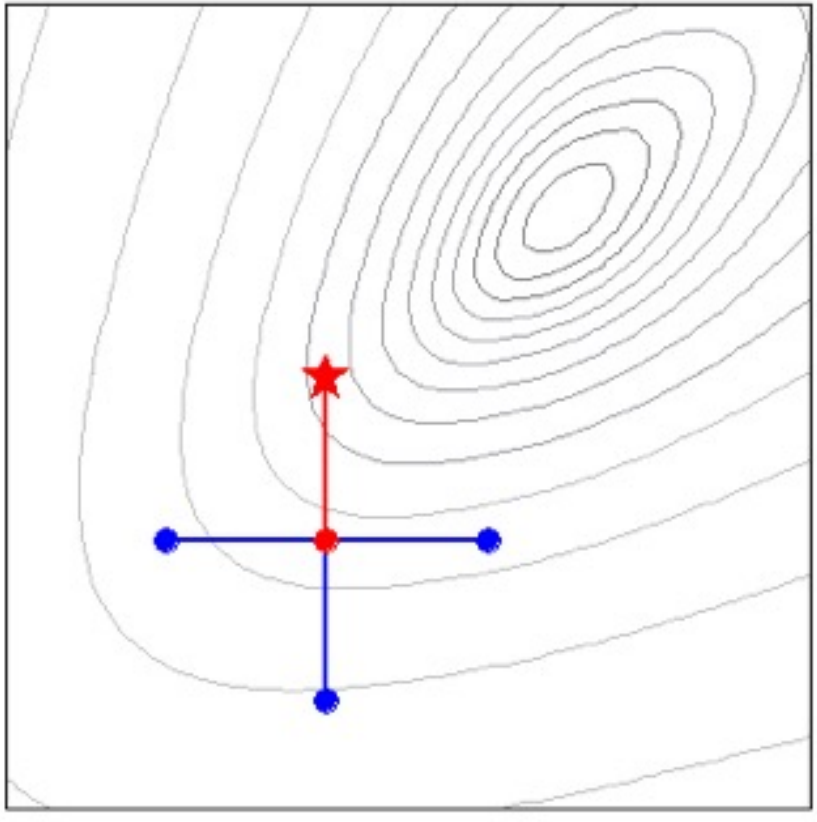
- Search phase can include user heuristics or surrogate functions
- Poll phase is more rigid, but needed to ensure convergence, i.e. sufficient descent directions.

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



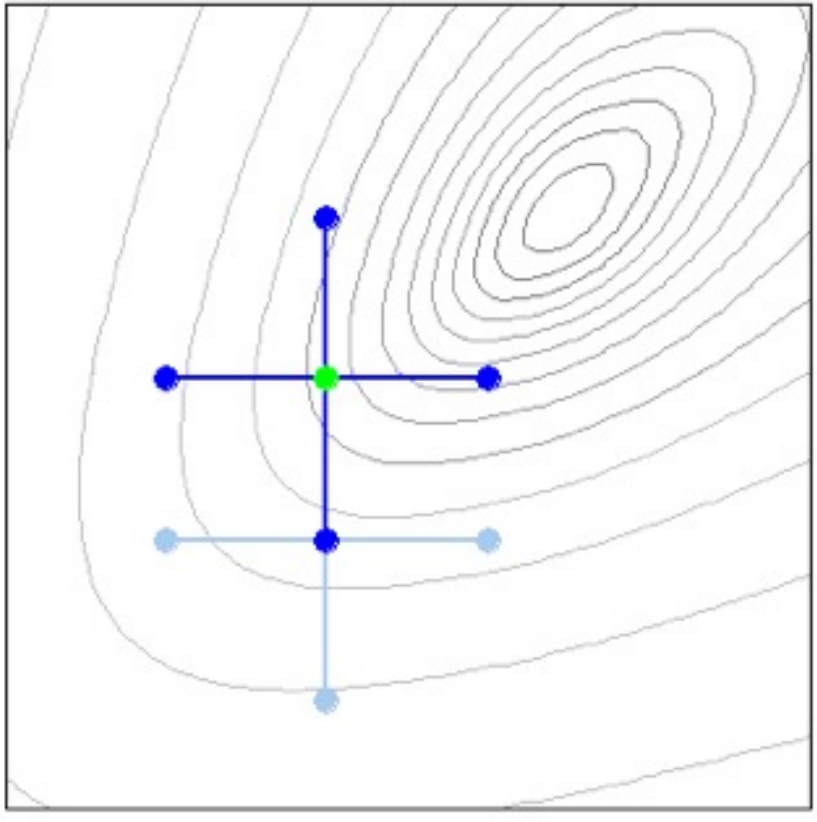
Special thanks to Tammy Kolda for this slide

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



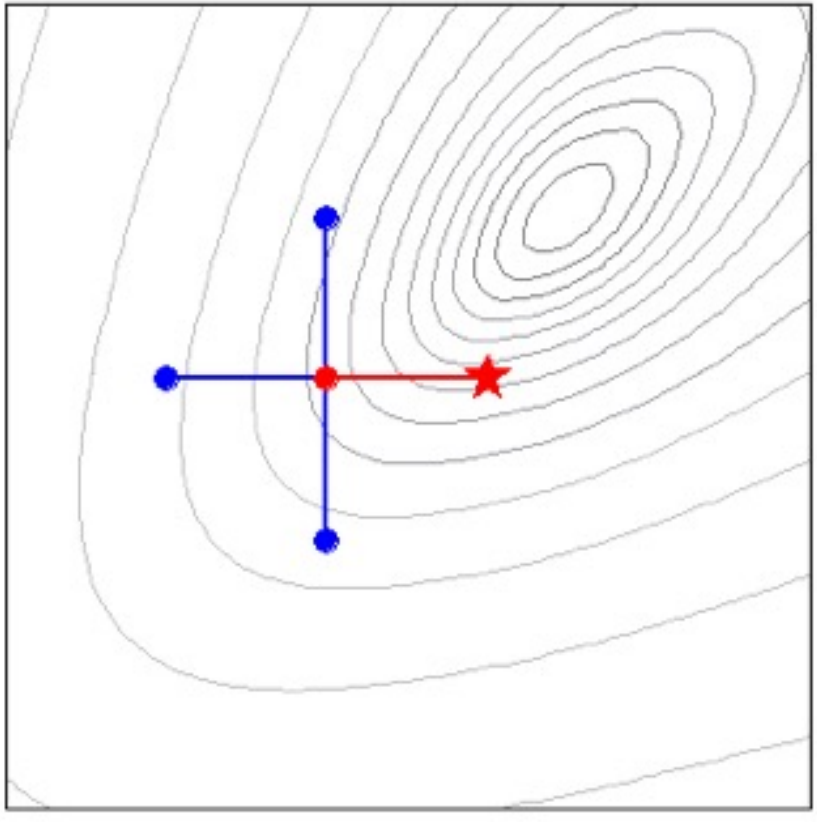
Special thanks to Tammy Kolda for this slide

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



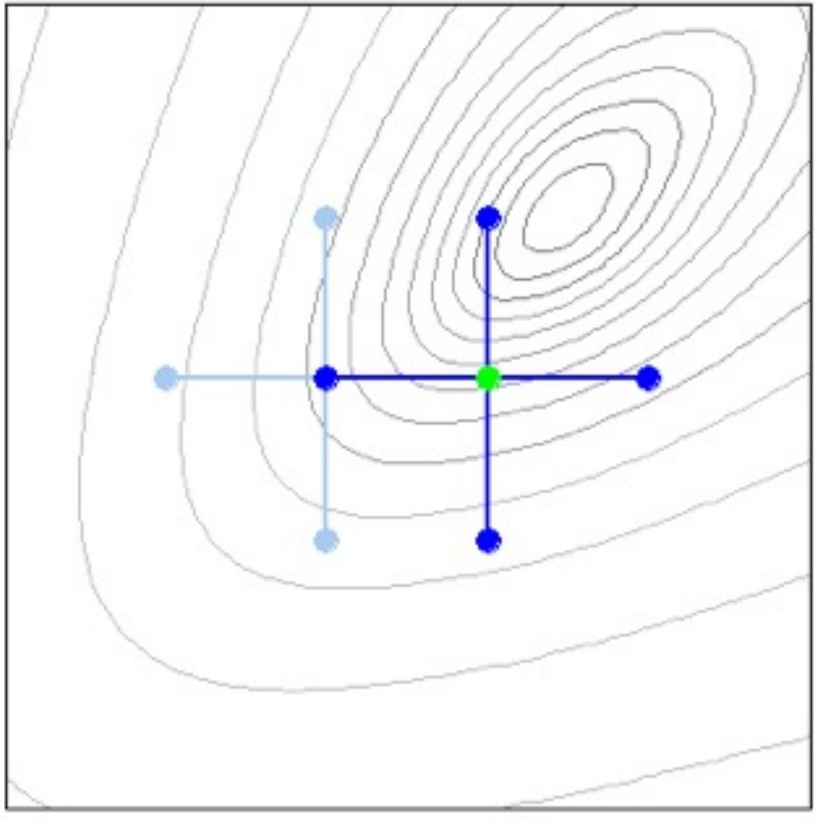
Special thanks to Tammy Kolda for this slide

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



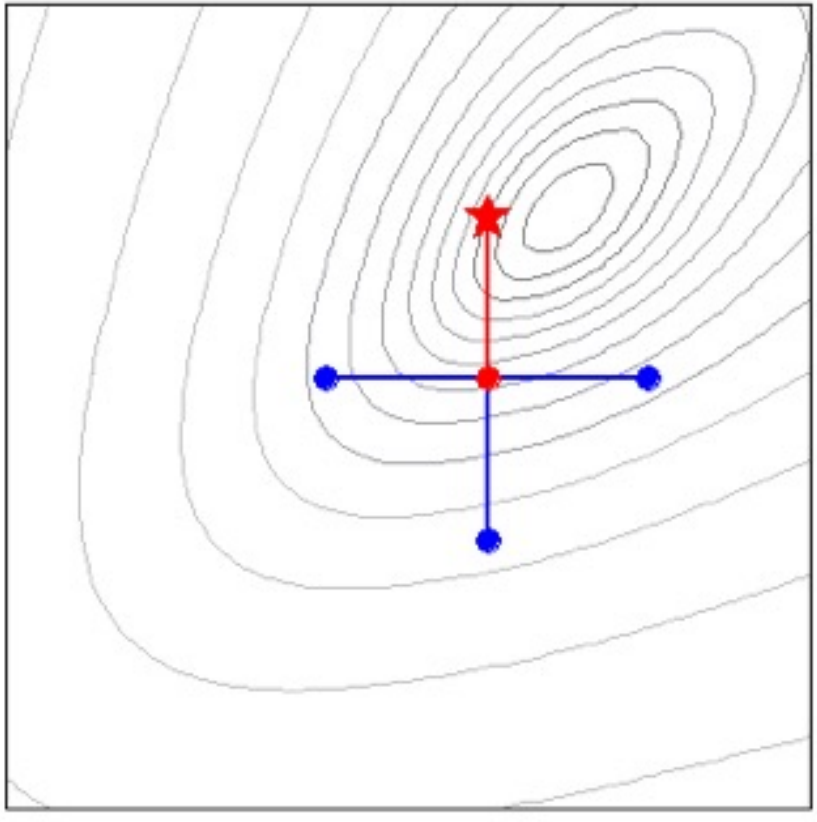
Special thanks to Tammy Kolda for this slide

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



Special thanks to Tammy Kolda for this slide

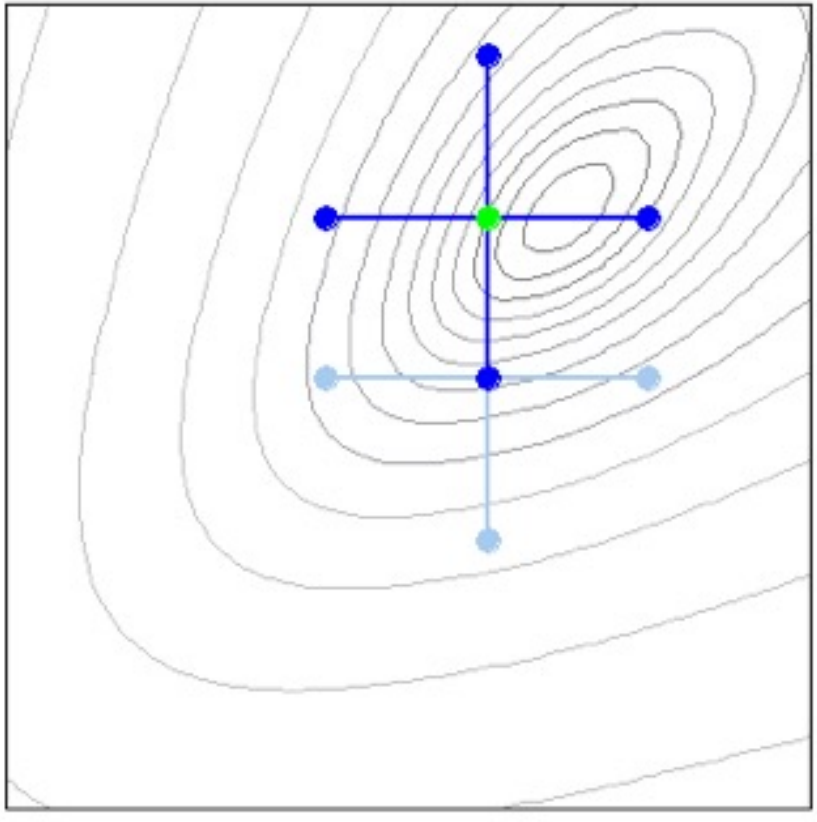
Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



Special thanks to Tammy Kolda for this slide

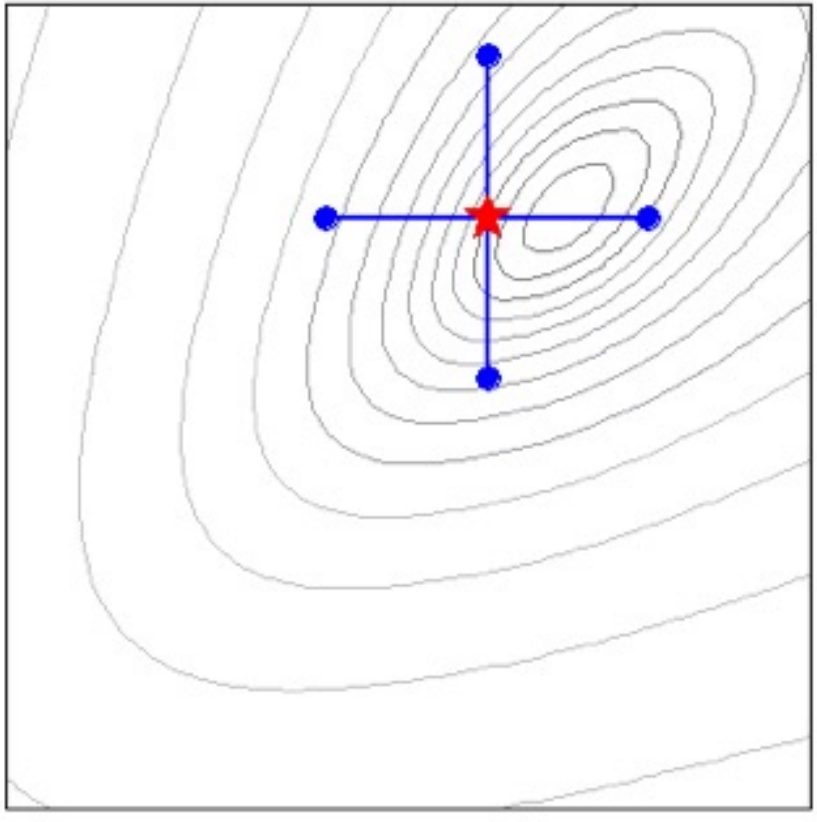
UCMERCED

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



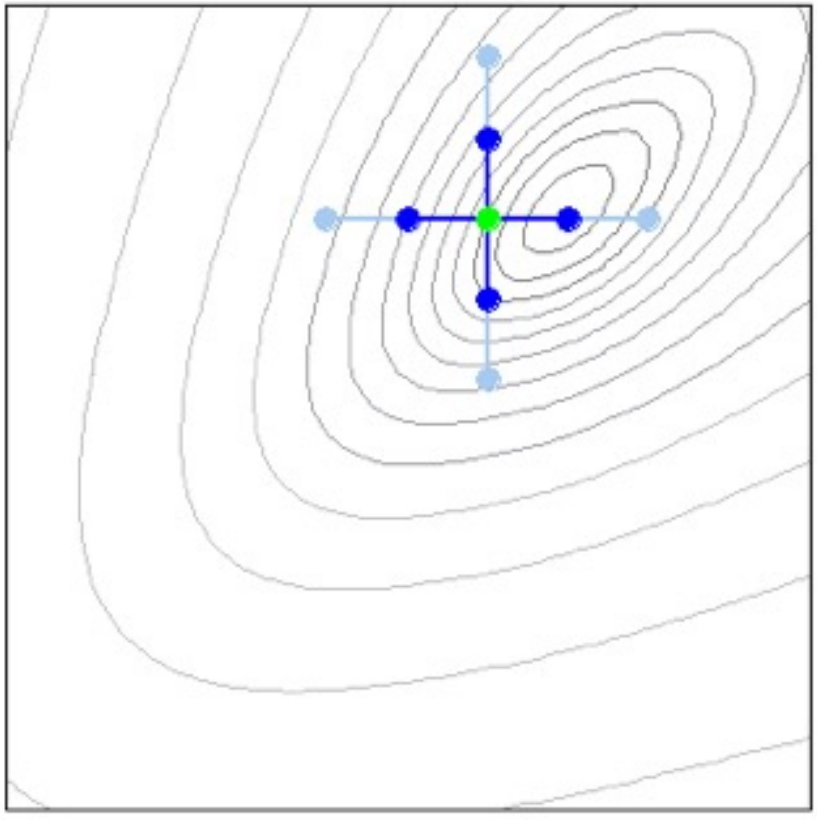
Special thanks to Tammy Kolda for this slide

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



Special thanks to Tammy Kolda for this slide

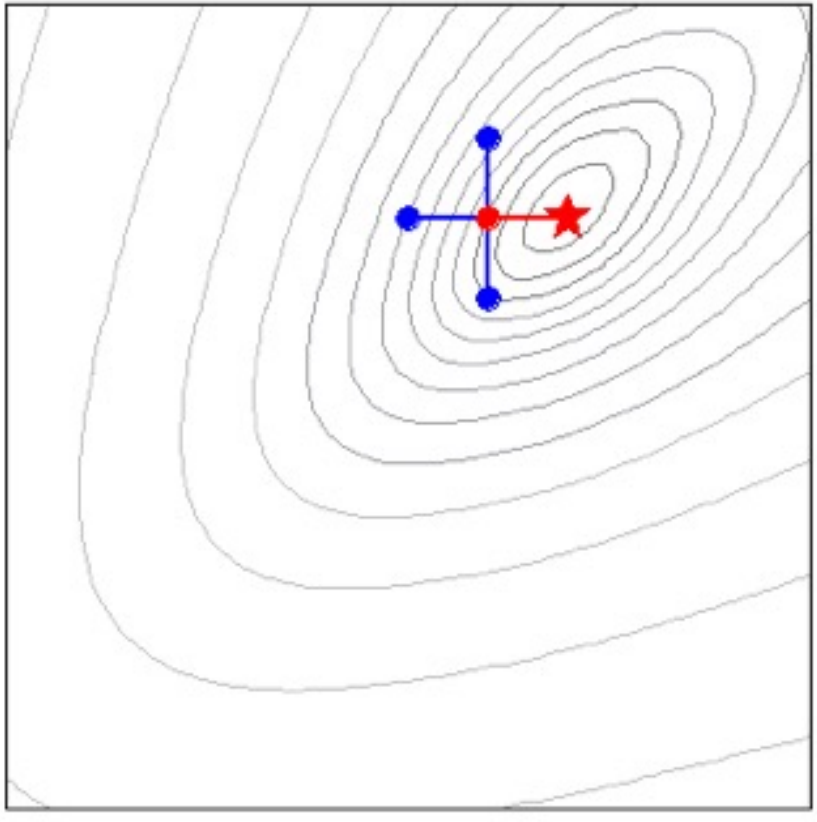
Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



Special thanks to Tammy Kolda for this slide

UCMERCED

Pattern Search $D_k = \{\pm e_1, \pm e_2\}$



Special thanks to Ta.....

UCMERCED

Global Convergence Properties, Part I

Generating Set Search (GSS) globalization strategies hinge on one key concept:

For all *unsuccessful* iterations

$$\lim_{k \to \infty} ||\Delta_k|| = 0$$

Various ways to produce such a sequence including

- sufficient decrease
- rational lattices
- moving grids

General assumptions mild including either f is bounded below or level sets are bounded.

Global Convergence Properties, Part II

If $f(x)$ is suitably smooth (continuously differentiable) can show that

- For unsuccessful iterations, $||\nabla f(x_k)||$ is bounded as a function of the step length Δ_k
- And since we can ensure $\liminf ||\Delta_k|| = 0$

Global Convergence

$$\liminf_{k \to \infty} ||\nabla f(x_k)|| = 0$$

N.B. We can prove stronger results under stricter conditions.

General Algorithm

Initialization: Given values Δ_0, x_0, M_0, P_0,

1 For $k = 0, \ldots,$ maxit do
2 Search: Evaluate f on a finite subset of trial points on a mesh M_k.
3 Poll: Evaluate f on the frame P_k.
4 If (successful)
5 $x_{k+1} = x_k + \Delta_k d_k$
6 Update Δ_k
7 End

- Search phase can include user heuristics or surrogate functions
- Poll phase is more rigid, but needed to ensure convergence, i.e. sufficient descent directions.

Taking advantage of LEED

- LEED
  - Multiple scattering model
  - I-V spectra computed repeatedly until best-fit structure is found
  - Computation time is proportional to the number of parameters
  - Most expensive of all methods
- KLEED
  - Assumes that electrons are only scattered once by the surface atoms.
  - Agrees well with experimental data as long as the incident angle is close to perpendicular
  - Surface unit cell size and symmetry can be determined, but not the exact positions
  - Compared to multiple scattering, KLEED I-V spectra can be computed order of magnitude faster

Idea: Use KLEED as a simplified physics surrogate

Building a better surrogate

- Define $\phi_A(x) = \phi_S + \phi_I$, where

ϕ_A =Additive surrogate

ϕ_S =Simplified physics surrogate, e.g. KLEED

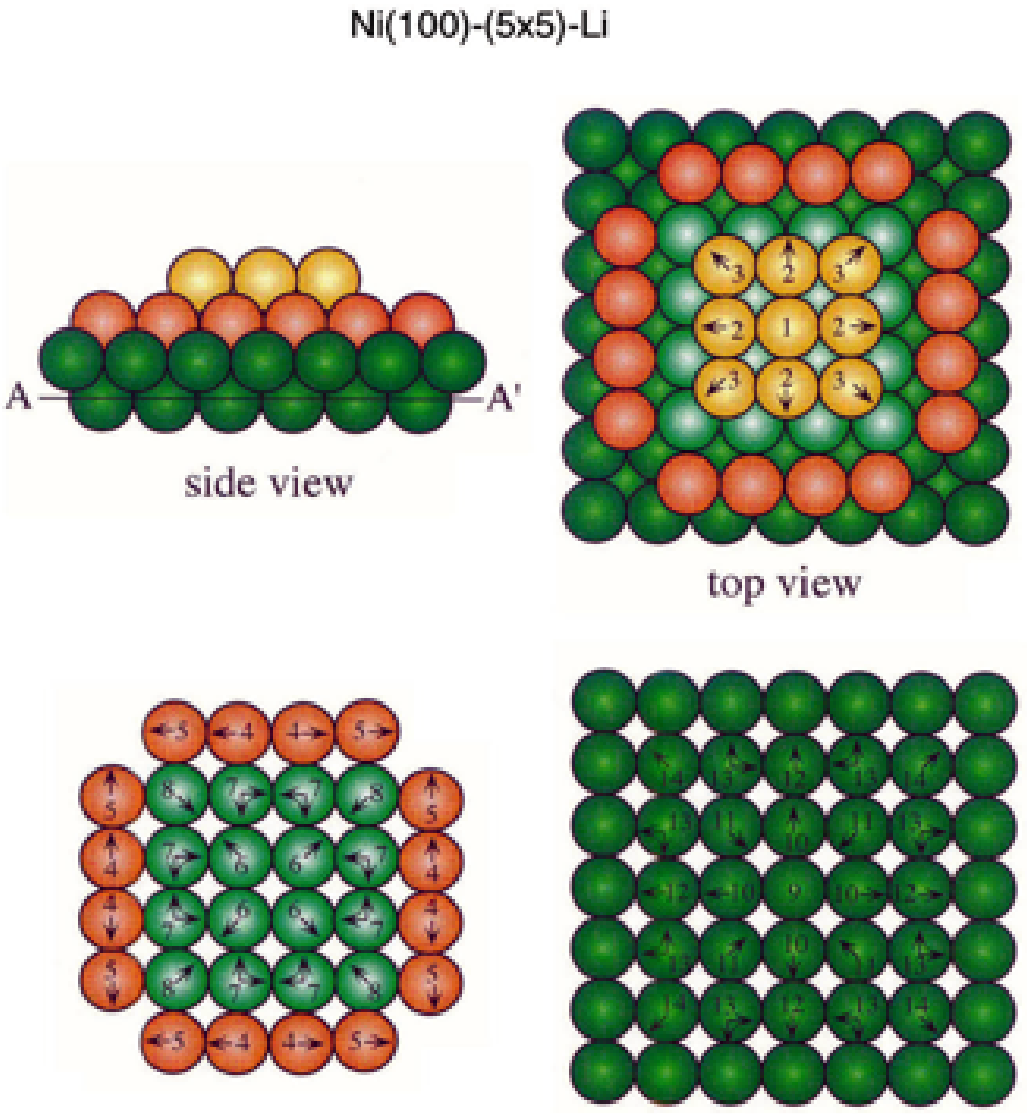
ϕ_I =Interpolatory surrogate

- Interpolatory surrogate designed to capture difference between simplified physics and "true" function values

Search:

```
1   if (first search step) {
2       Initialize φ_I with Latin Hypercube sample.
3   else {
4       recalibrate φ_I with DACE model of φ_S − f.
5       }
6   Construct φ_A = φ_S + φ_I
7   Solve min φ_A
```

Test problem

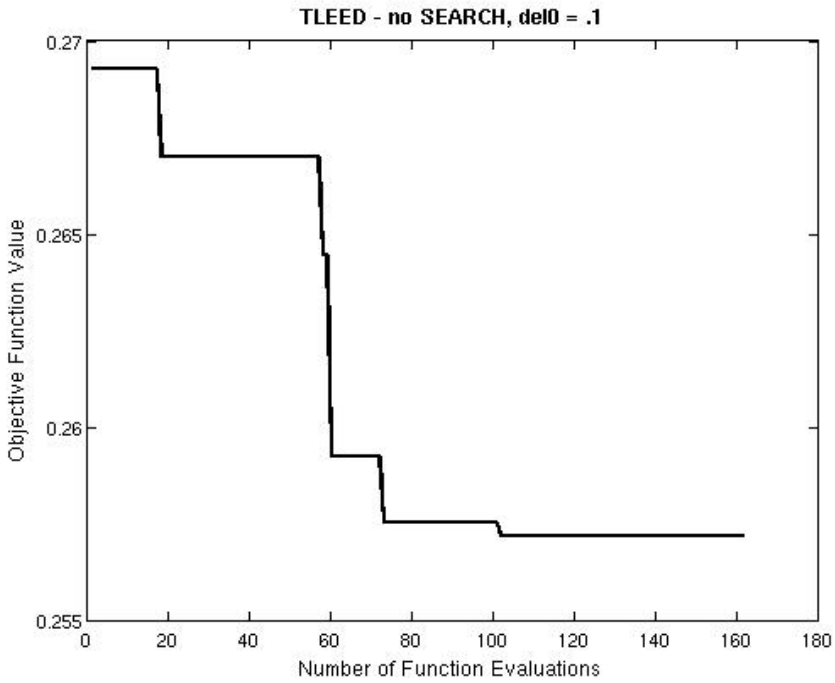


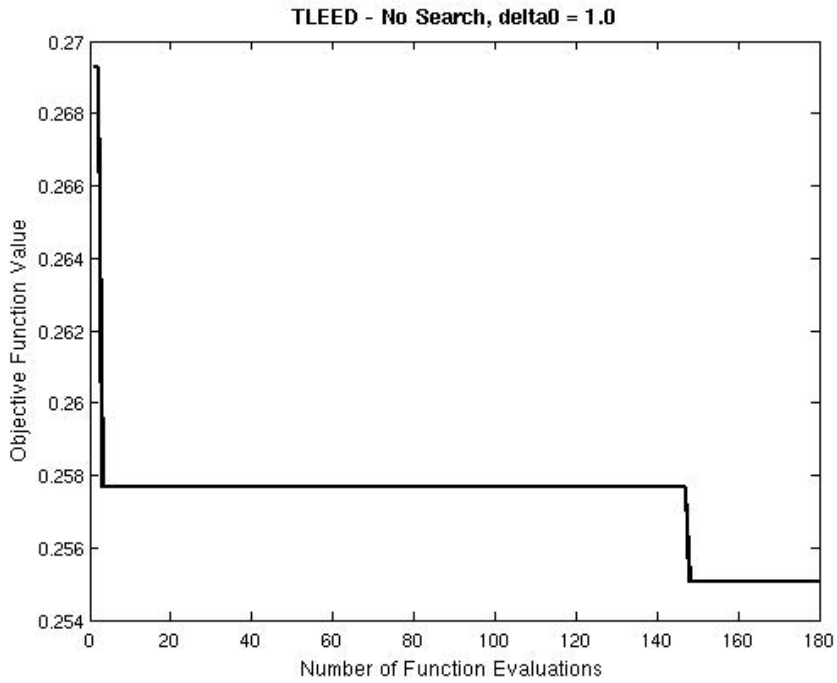
- Model 31 from set of model problem using three layers
  - 14 atoms
  - 14 categorical variables
  - 42 continuous variables
- Additional constraint added so that positions of atoms are constrained to lie within a box
- Used NOMADm (Abramson, Audet, Dennis, Le Digabel, Tribes)

Test cases

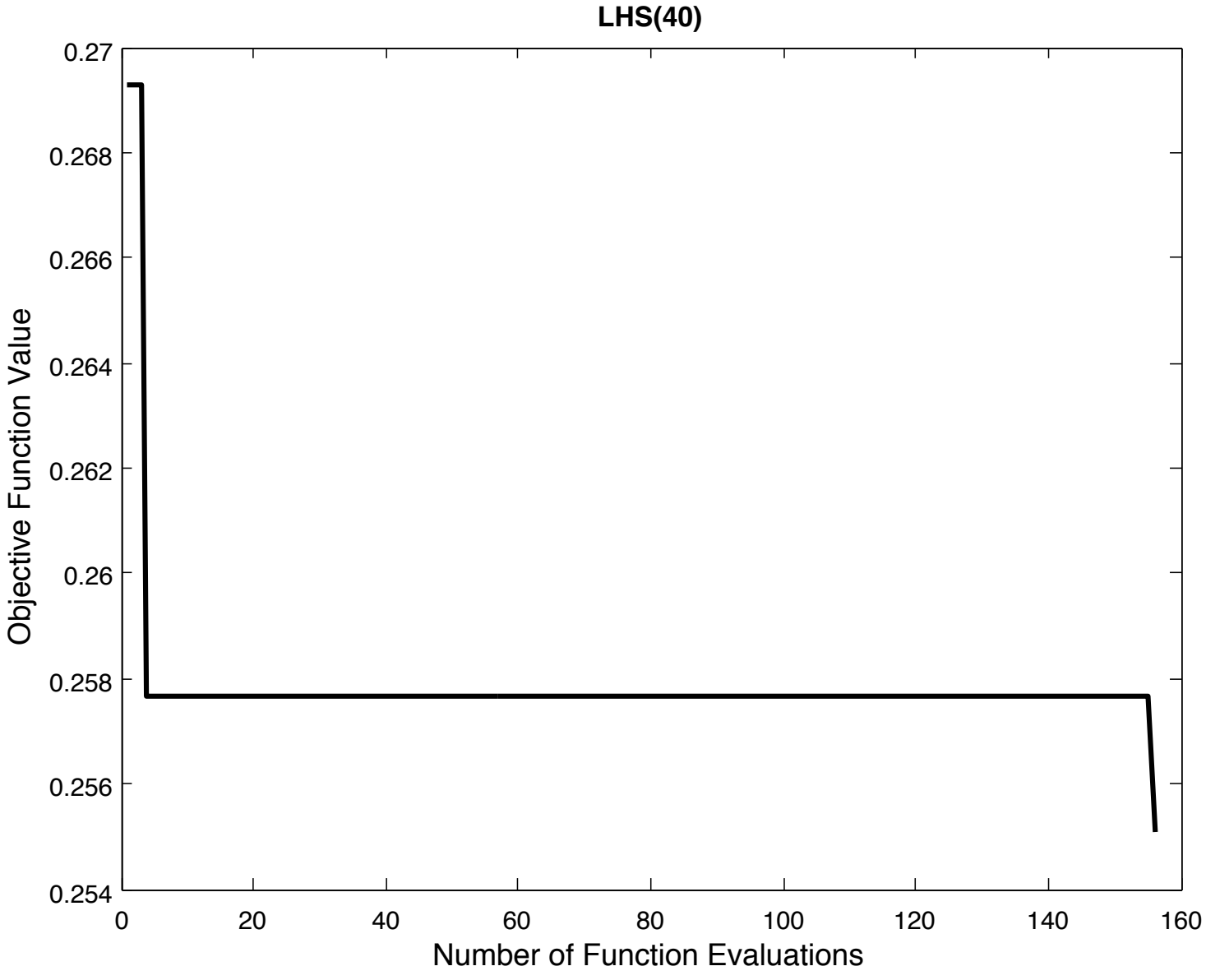
- Start with best known feasible point
- Continuous variables only
- Three different approaches
  - No search step
  - LHS search
  - Simplified physics surrogate (DACE model)

Optimization of continuous variables using no search phase

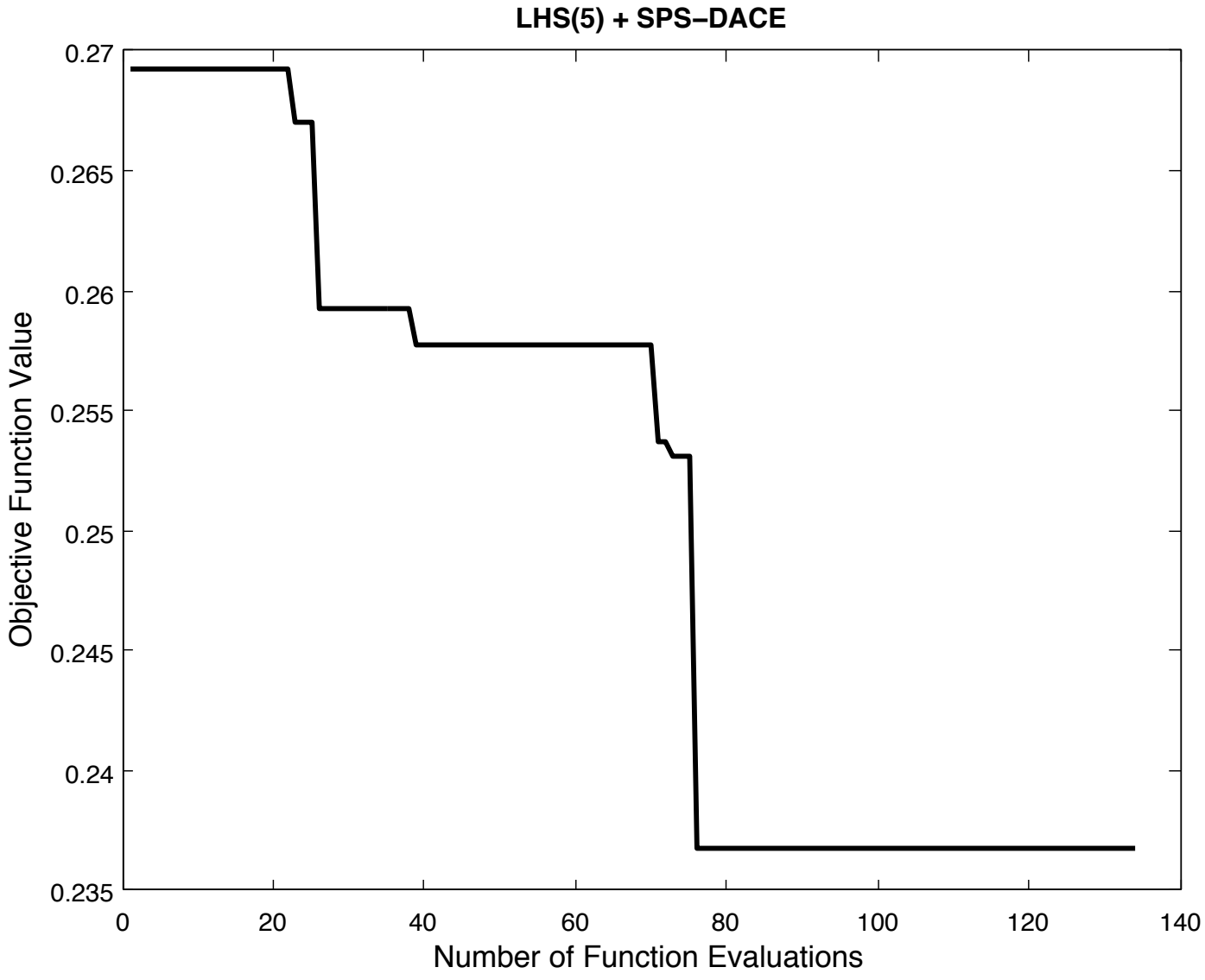




Optimization using LHS with 40 points



Optimization using additive surrogate, $\Delta_0 = 0.1$



Summary of numerical results

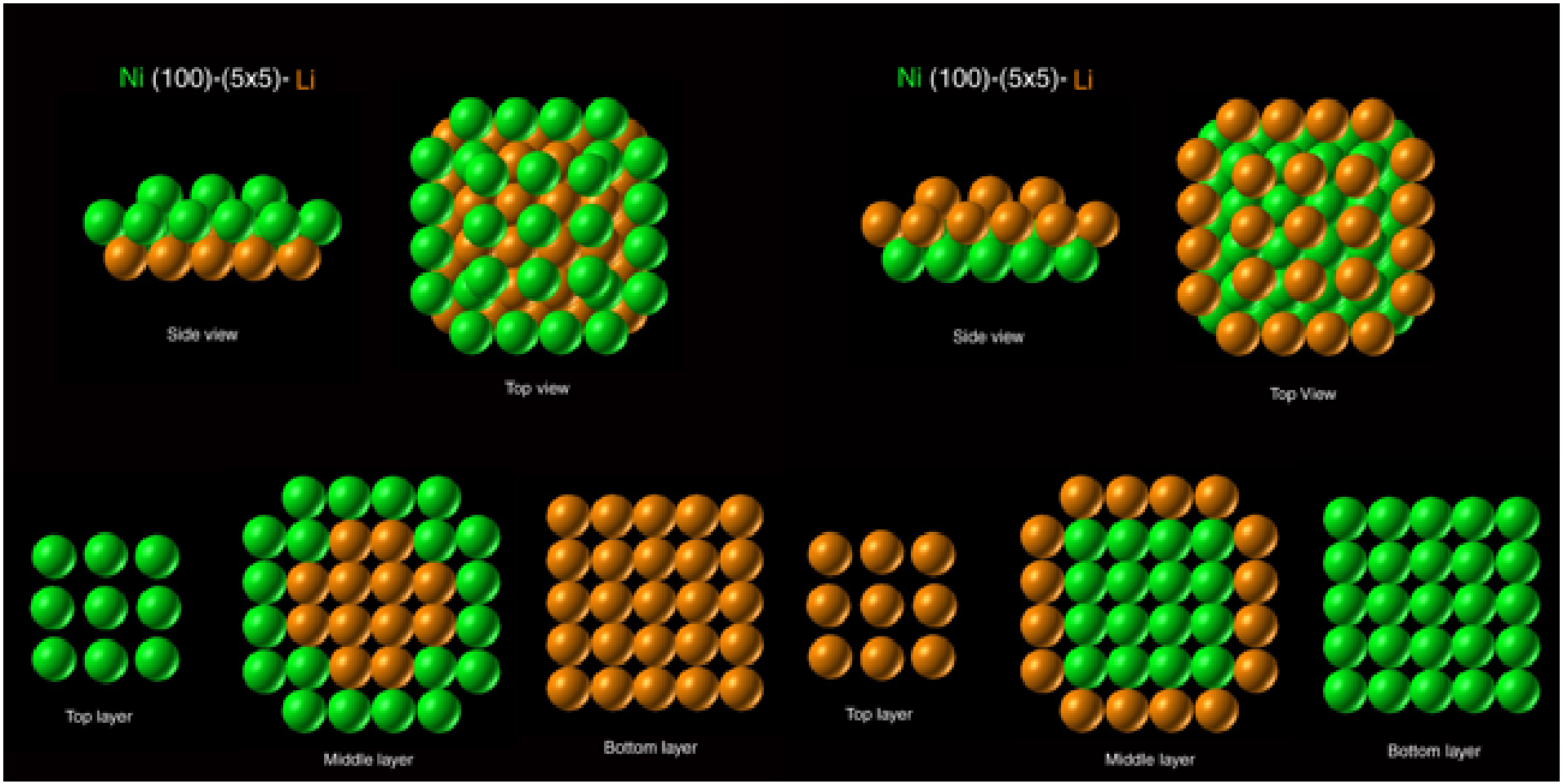
Method	LHS	$f(x^*)$	fevals
No search	0	0.2551	180
LHS	40	0.2551	160
SPS+DACE	15	0.2543	180
SPS+DACE	5	0.2354	135

Summary

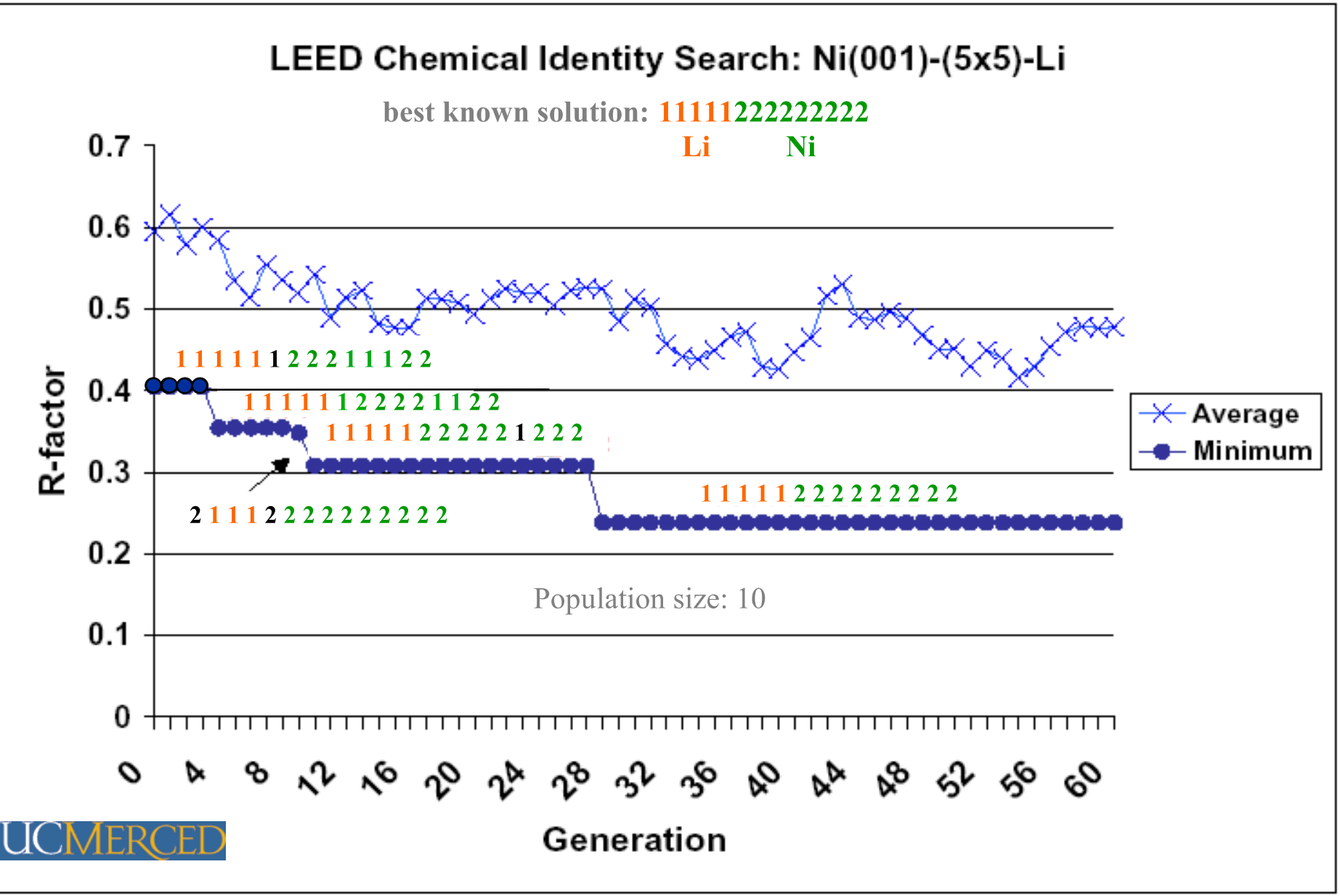
- Simulation-Based Optimization is an important and rapidly growing area of research
- Many standard assumptions do not apply for this class of problems – great opportunity for new ideas and research
- Presented one example that used a combination of adapting old methods and taking advantage of structure
  - Total number of function evaluations decreased by about 20% per model
  - Room for improvement and new ideas, e.g. alternate interpolatory surrogates
  - Another direction is solving the full mixed variable problem

Questions?

Future Work – Chemical Identity Search

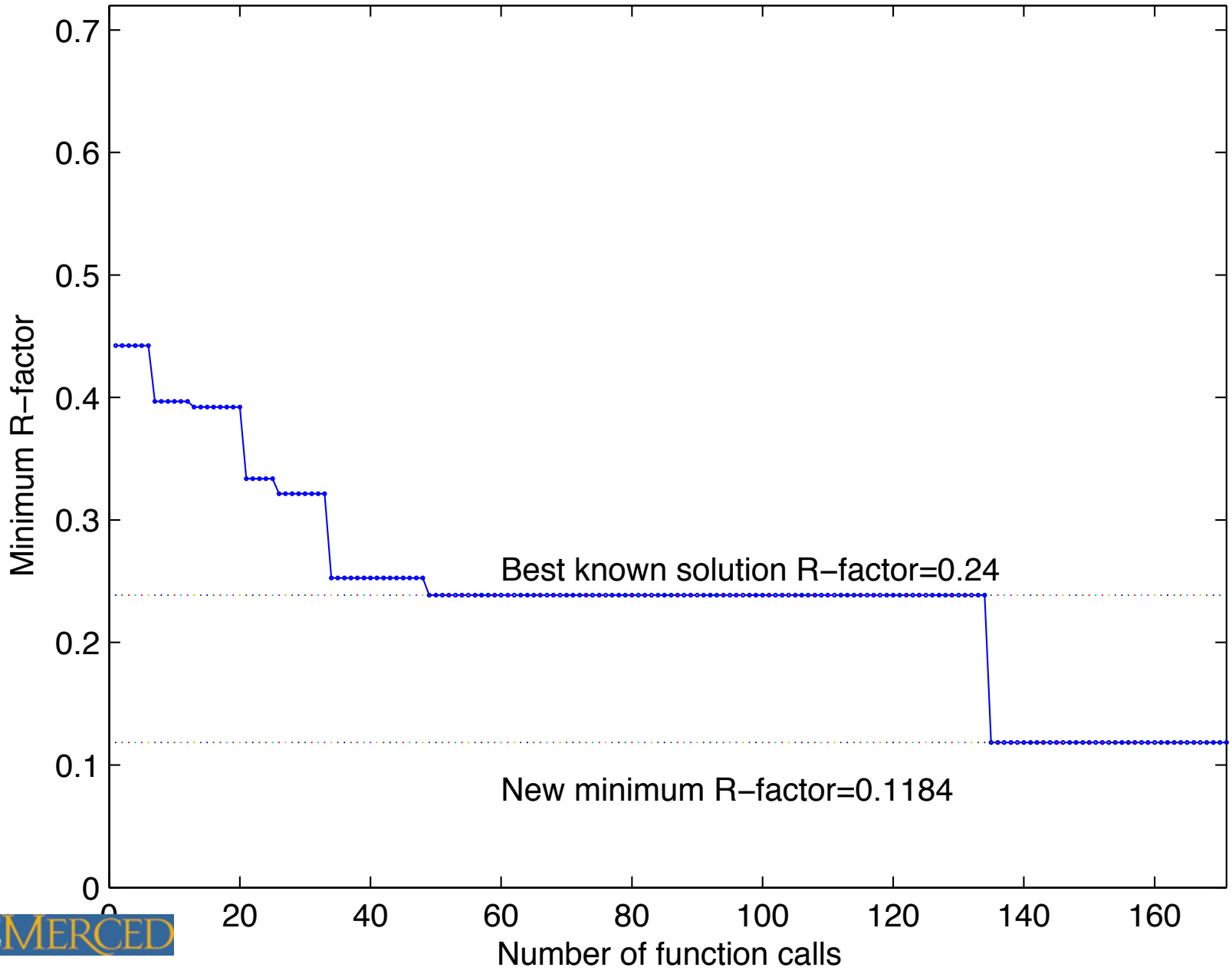


GA with Chemical Identity



UCMERCED

Pattern Search: Chemical Identity + Cont. Vars



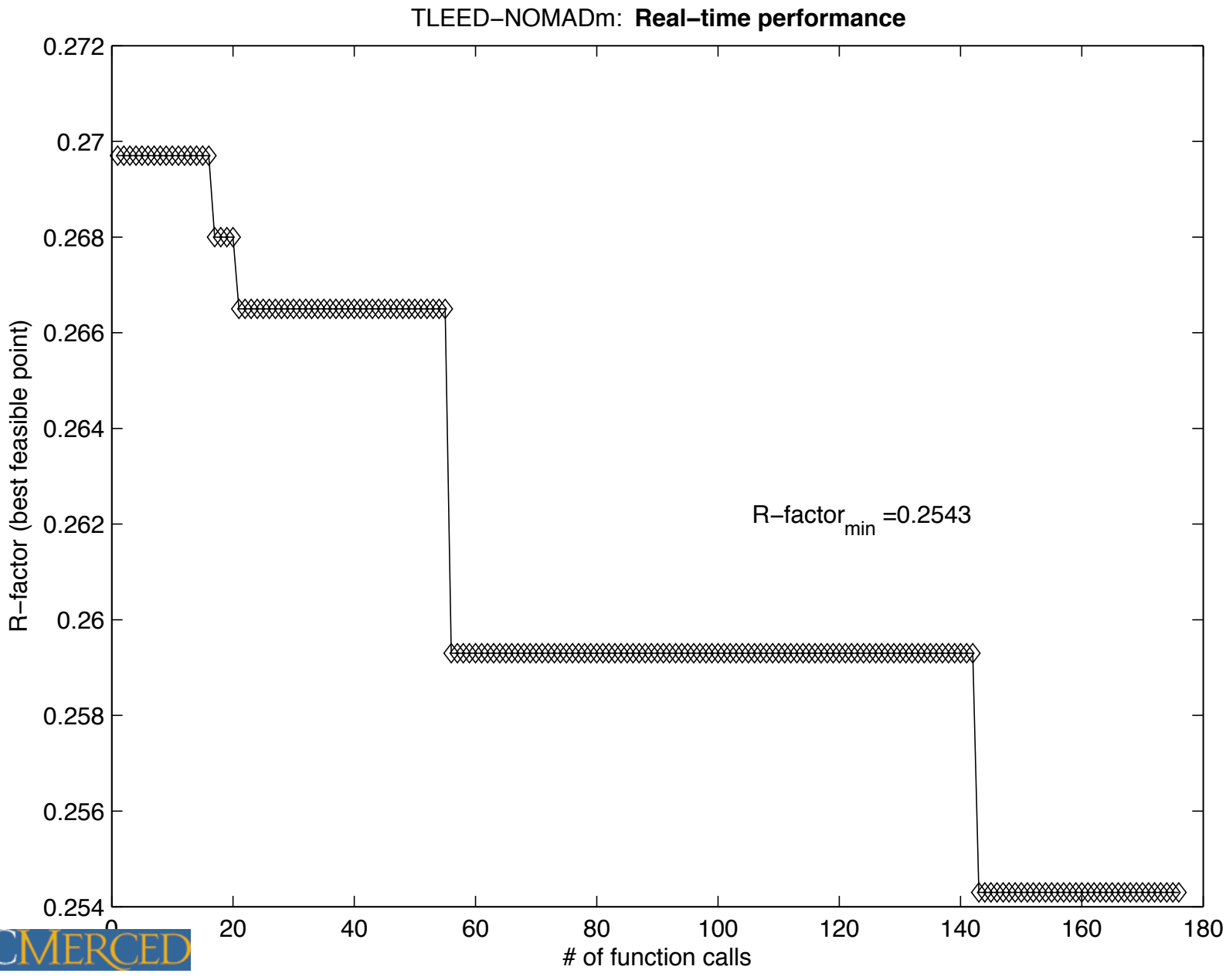
Pendry Reliability-factor (1980)

$$
\begin{aligned}
R &= \sum_{i,g}(Y_{gth} - Y_{gexp})^2 / \sum_{i,g}(Y_{gth}^2 + Y_{gexp}^2) \\
Y(E) &= L/(1 + L{V_{oi}}^2) \\
L(E) &= I'/I \\
L &\approx \sum_j \frac{-2(E - E_j)}{(E - E_j)^2 + V_{oi}^2}
\end{aligned}
$$

Pendry R-factor

- LEED curves consist for the main part of a series of Lorentzian peaks: $I \approx \sum a_j/(E - E_j)^2 + V_{oi}^2$
- Their widths are dictated by the imaginary part of the electron self-energy (optical potential): $\Delta E = 2|V_{oi}|$
- Pendry R-factor emphasizes positions of the maximum and minimum rather than the heights of the intensities
- Ideal agreement corresponds to $R = 0$; uncorrelated spectra yields $R = 1$.

Optimization using additive surrogate, $\Delta_0 = 1.0$



Kinematic LEED

- KLEED assumes that electrons are only scattered once by the surface atoms.
- Agrees well with experimental data as long as the incident angle is close to perpendicular.
- Surface unit cell size and symmetry can be determined, but not the exact positions.
- Compared to multiple scattering which takes several minutes to compute, I-V spectra from KLEED can be computed in a few seconds.

Idea: Use KLEED as a simplified physics surrogate

Pendry Reliability-factor (1980)

- IV curves consist for the main part of a series of Lorentzian peaks (I is Intensity):

$$I \approx \sum \frac{a_j}{(E - E_j)^2 + V_{oi}^2}$$

- Pendry R-factor emphasizes positions of the peaks rather than the heights of the intensities
- Ideal agreement corresponds to $R = 0$; uncorrelated spectra yields $R = 1$.

$$\begin{aligned} L(E) &= I'/I \\ L &\approx \sum_j \frac{-2(E - E_j)}{(E - E_j)^2 + V_{oi}^2} \\ Y(E) &= L/[1 + (LV_{oi})^2] \\ R &= \frac{\sum_{i,g}(Y_{gth} - Y_{gexp})^2}{\sum_{i,g}(Y_{gth}^2 + Y_{gexp}^2)} \end{aligned}$$