Challenges for AI in Computational Sustainability

Carla P. Gomes
Institute for Computational Sustainability
Cornell University

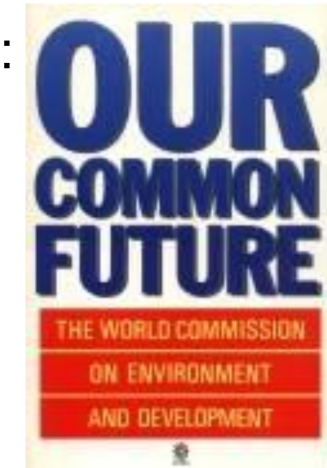
Support:

Expeditions in Computing (CISE)

AAAI
Atlanta
July 2010

Sustainability and Sustainable Development

The 1987 UN report, “Our Common Future” (Brundtland Report):



- Raised serious concerns about the State of the Planet.
- Introduced the notion of sustainability and sustainable development:

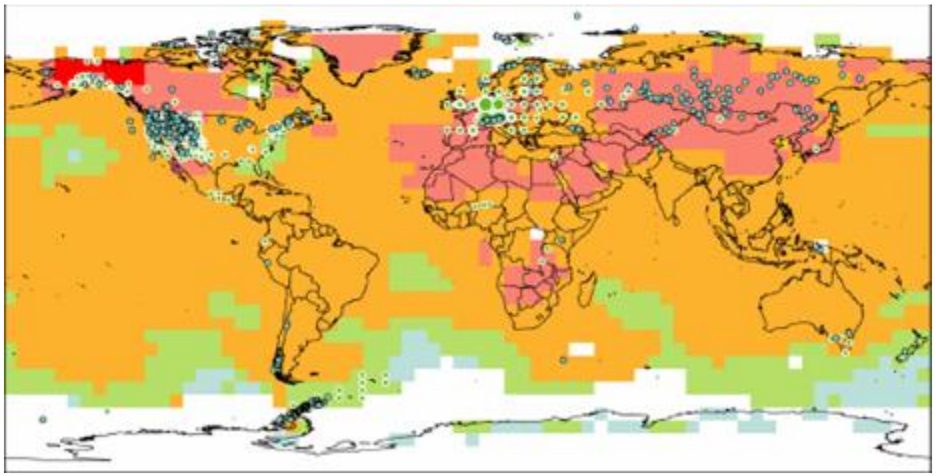
Sustainable Development: “development that meets the needs of the present without compromising the ability of future generations to meet their needs.”

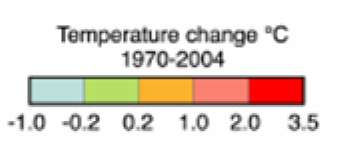
UN World Commission on Environment and Development,1987.

Gro Brundtland
Norwegian Prime Minister
Chair of WCED

Follow-Up Reports:
Intergovernmental Panel on Climate Change (IPCC 07) Global Environment Outlook Report (GEO 07)

"There are no major issues raised in Our Common Future for which the foreseeable trends are favourable."





Global Warming

Erosion of Biodiversity

Examples:

- **At the current rates of human destruction of natural ecosystems, 50% of all species of life on earth will be extinct in 100 years.**

+130 countries

WMO

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC)

UNEP

[Nobel Prize with Gore 2007]

Poor Management of our Natural Resources

Pollution





Habitat Loss and Fragmentation



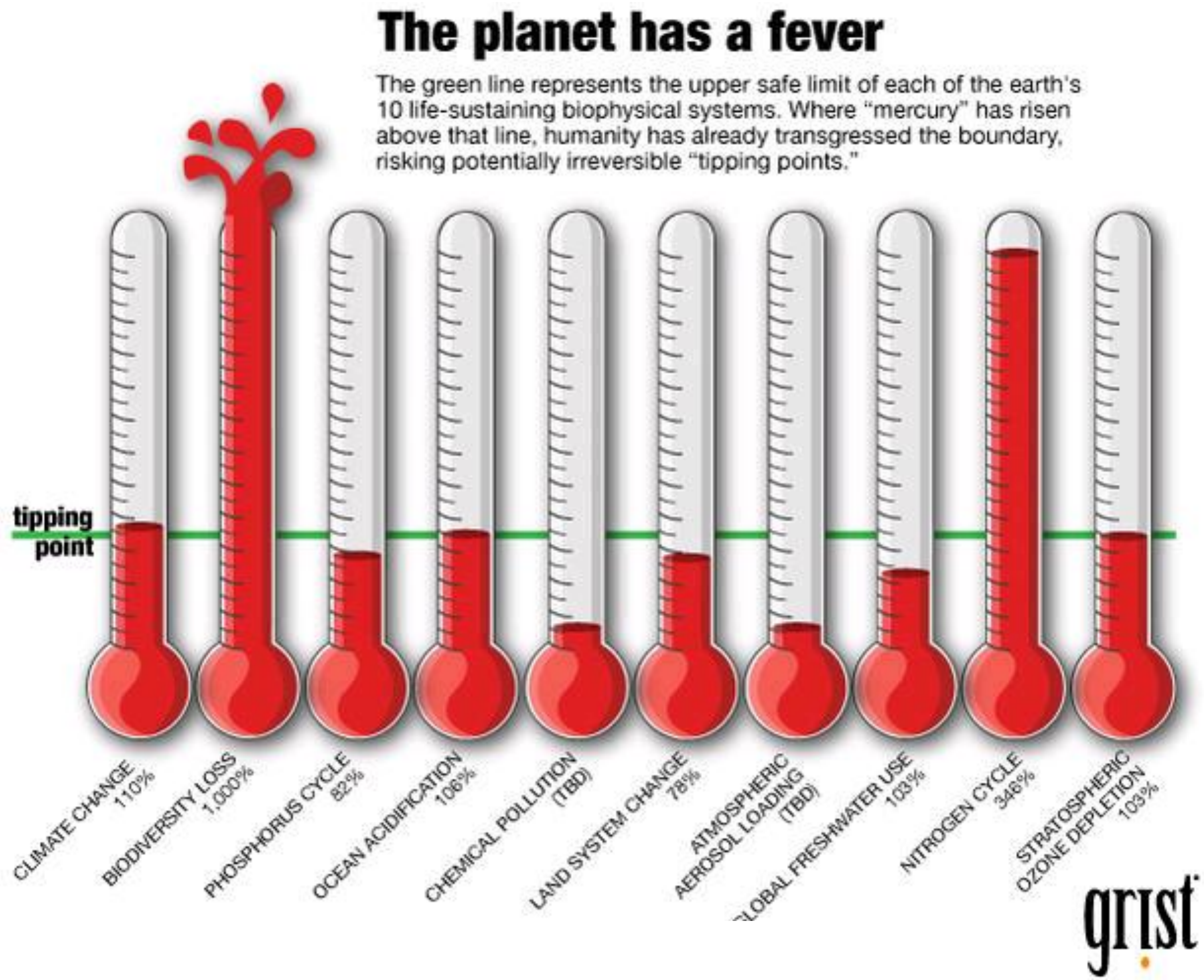


Over-Harvesting






Safe Operating Boundaries: Crucial Biophysical Systems



Source:

***Planetary Boundaries: A Safe Operating Space for Humanity*, Nature, 2009**

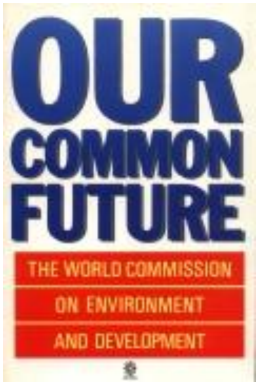
Sustainability: Not Only About the Environment

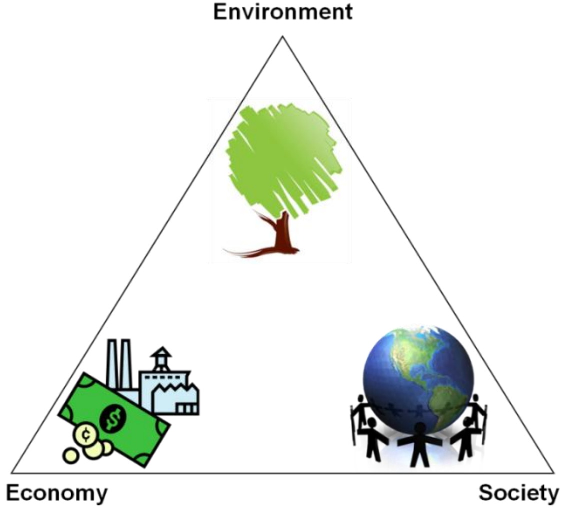
Our Common Future recognized that environmental, economic and social issues are interlinked.

→The economy only exists in the context of a society, and both society and economic activity are constrained by the earth's natural systems.

→A secure future depends upon the health of all 3 systems (environment, society, economy).

Sustainable Development encompasses balancing environmental, economic, and societal needs.





Policies for Sustainable Development

- The UN reports stressed the urgency of policies for sustainable development.

- Key issues pertaining to the design of policies for sustainable development translate into decision, optimization, and learning problems concerning the management of our natural resources

- Often such problems involve **significant computational challenges** that fall into the realm of computing and information science but are **not studied by computer scientists**; e.g.,

  – **Constraint Reasoning and optimization, machine learning, and data mining, dynamical systems etc.**

Vision

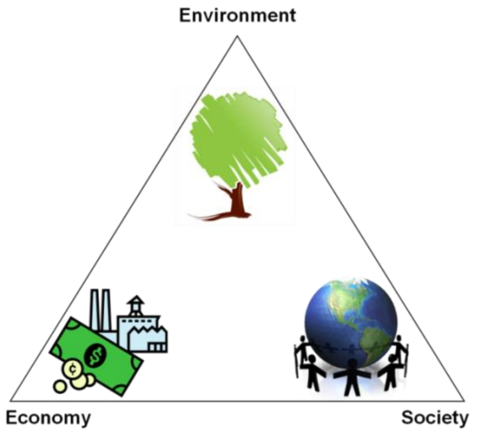
Computer scientists can — and should — play a key role in increasing the efficiency and effectiveness of the way we manage and allocate our natural resources, while enriching and transforming Computer Science.

Why we need critical mass in the new field of
Computational Sustainability!!!

Computational Sustainability: attempt at a definition

New interdisciplinary field that aims to apply techniques from computer science, and related fields (e.g., information science, operations research, applied mathematics, and statistics) for Sustainable Development.

Sustainable Development encompasses *balancing environmental, economic, and societal needs.*



Computational Sustainability
attempt at a definition

Wide interdisciplinary field , encompassing disciplines as diverse as economics, sociology, environmental sciences and engineering , biology, crop and soil science, meteorology and atmospheric science.

Focus:

Develop computational models, methods and tools for decision making for a broad range of sustainability related applications:

from decision making and policy analysis concerning the management and allocation of natural resources

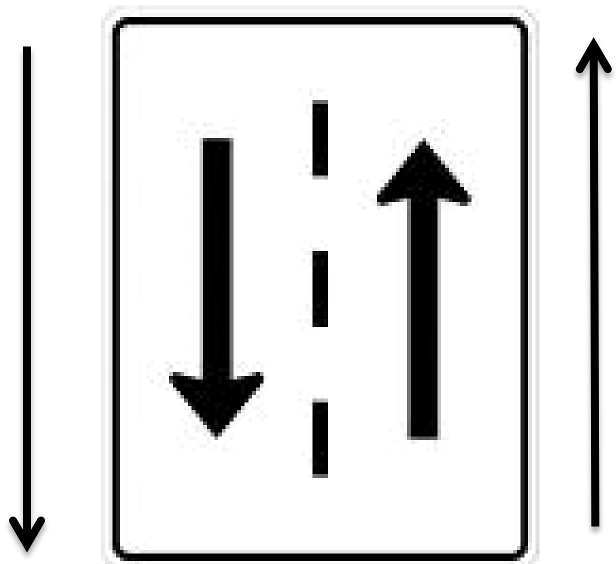
to the design of new sustainable techniques, practices and products.

Key challenge: to effectively and efficiently establish interdisciplinary collaborations – the level of interconnectedness of social, economic, and environmental issues makes it really challenging!

Sustainability fields

New challenging applications
for Computer Science

New methodologies
In Computer Science



Computational Thinking that will provide new insights into sustainability problems:

New methodologies
In Sustainability fields

Computer science and related fields

Analogy with **Computational Biology**

Outline

- Computational Sustainability Research @ The Institute for Computational Sustainability (ICS)
  - Research team
  - A sample of Interdisciplinary Research Projects (IRPs)
  - Main computational themes
- Other Activities at the Institute for Computational Sustainability
- Building a community in Computational Sustainability
- Conclusions

Thanks! The Institute for Computational Sustainability (ICS) Research Team

ICS members and collaborators for their many contributions towards the development of a vision for Computational Sustainability

Sabharwal
CS
Cornell

Walker
Bio & Env. Eng.
Cornell

Bento
Res. & Env. Economics
Cornell

Shmoys
CS & OR
Cornell

Strogatz
Appl. Math
Cornell

Mahowald
Earth & Atmos. Sci.
Cornell

Gomes
CS
Cornell

Dietterich
CS
OSU

Zeeman
Appl. Math
Bowdoin

Yakubu
Appl. Math
Howard

Montgomery
Res. & Env. Economics
OSU

Hopcroft
CS
Cornell

Albers
Res. & Env. Econo.
OSU

Chavarria
HPC
PNNL

A Multi-institutional, Multidisciplinary Research Team

6 Institutions, 7 colleges, 13 departments

Cooch
Natural Resources
Cornell

Ellner
Ecology & Evol. Bio.
Cornell

Wong
CS
OSU

Conrad
Res. and Env. Economics
Cornell

Amundsen
Conservation Planning
Cons. Fund

Selman
CS
Cornell

DiSalvo
Chemistry
Cornell

Sofia
Biology

Barrett
Res. & Env. Economics
Cornell

Guckenheimer
Math
Cornell

Rosenberg
Conserva-tion Biology

McDonald

Damoulas
CS
Cornell

Support: NSF Expeditions in Computing (CISE)

Cornell University Founded A.D. 1865
Bowdoin
OSU
HOWARD UNIVERSITY
THE CONSERVATION FUND
America's Partner in Conservation

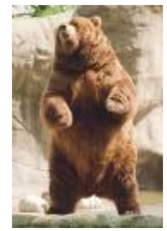
29 graduate students
24 undergrad. students

Sample of Projects @ ICS

I Environment

Conservation and Biodiversity

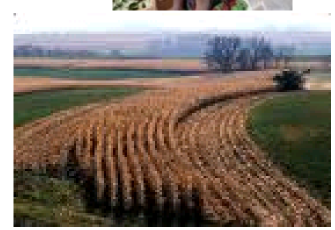
E.g., Wildlife corridors, species distributions, bird migration

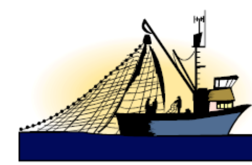


II Economy

Socio-economic systems

E.g. Poverty mapping and poverty reduction, harvesting policies, agricultural systems

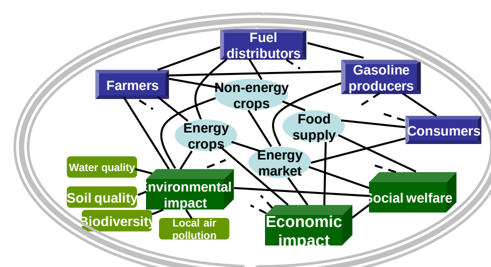




III Energy

Renewable Energy

E.g. Material discovery and Biofuels



Conservation and Biodiversity

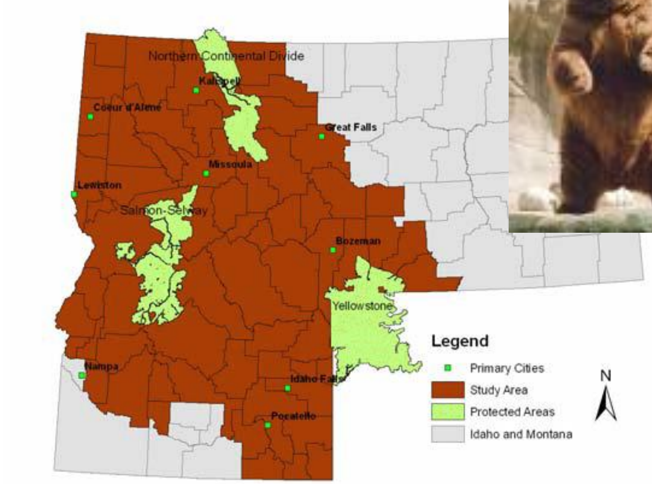
Conservation and Biodiversity : Wildlife Corridors

Wildlife Corridors link core biological areas, allowing animal movement between areas.

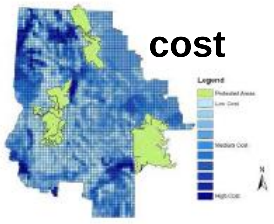
Typically: low budgets to implement corridors.

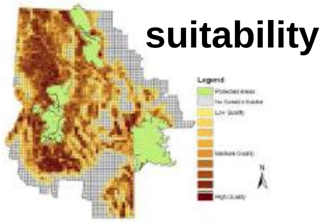
Example:

Goal: preserve grizzly bear populations in the U.S. Northern Rockies by creating wildlife corridors connecting 3 reserves:



Yellowstone National Park
Glacier Park / Northern Continental Divide
Salmon-Selway Ecosystem

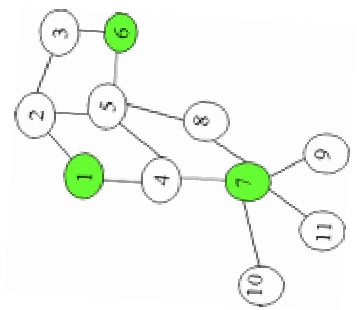


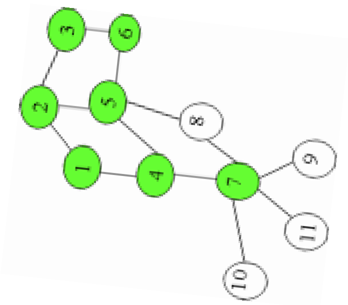


Conservation and Biodiversity: Wildlife Corridors
Challenges in Constraint Reasoning and Optimization

Wildlife corridor design

Computational problem → **Connection Sub-graph Problem**





Connection Sub-Graph - NP-Hard

Worst Case Result --- Real-world problems possess hidden structure that can be exploited allowing scaling up of solutions.

Connection Sub-graph Problem

Given a graph G with a set of reserves:

Find a sub-graph of G that:

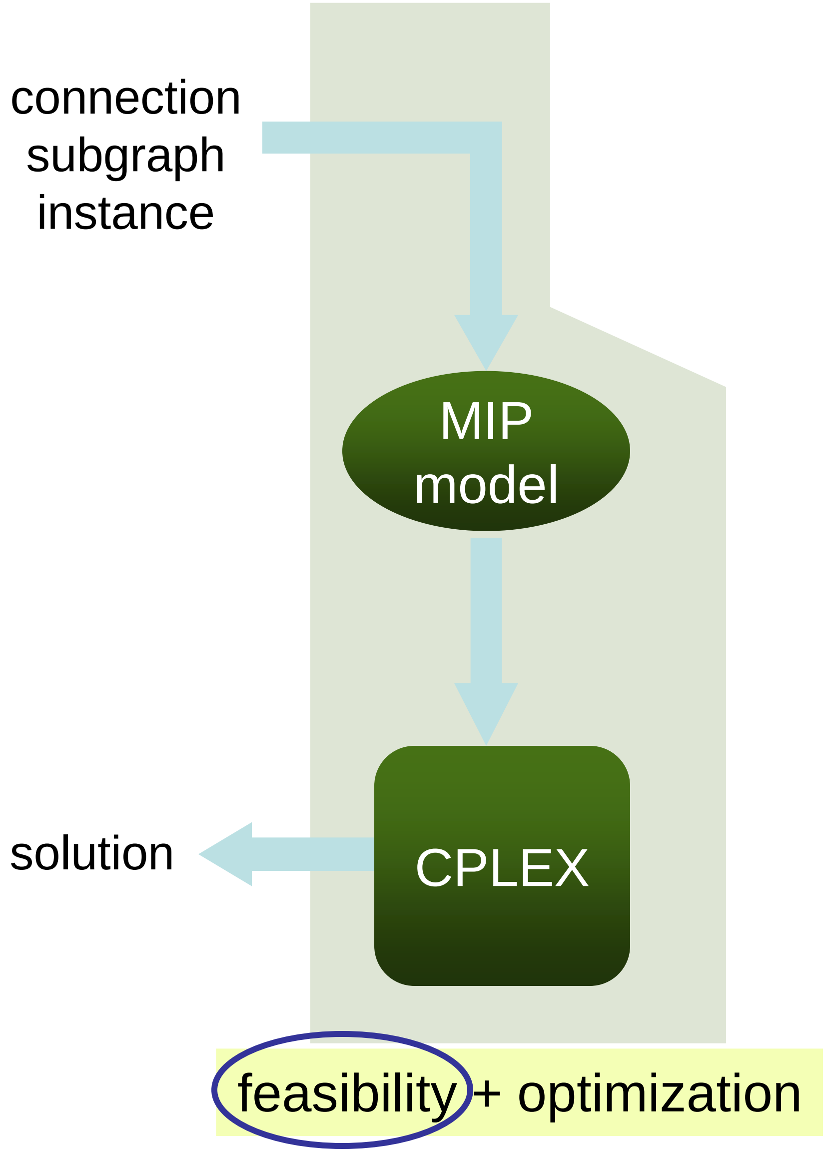
contains the reserves;
is fully connected;
with cost below a given budget;

and
with maximum utility

Interdisciplinary Research Project:

Conrad, Dilkina, Gomes, van Hoeve, Sabharwal, Sutter; 2007-2010

Solving the Connection Sub-Graph Problem: Standard Mixed Integer Programming (MIP) Approach



- MIP model based on network flow
- Revealed interesting tradeoffs between testing for infeasibility and optimization

Problem?

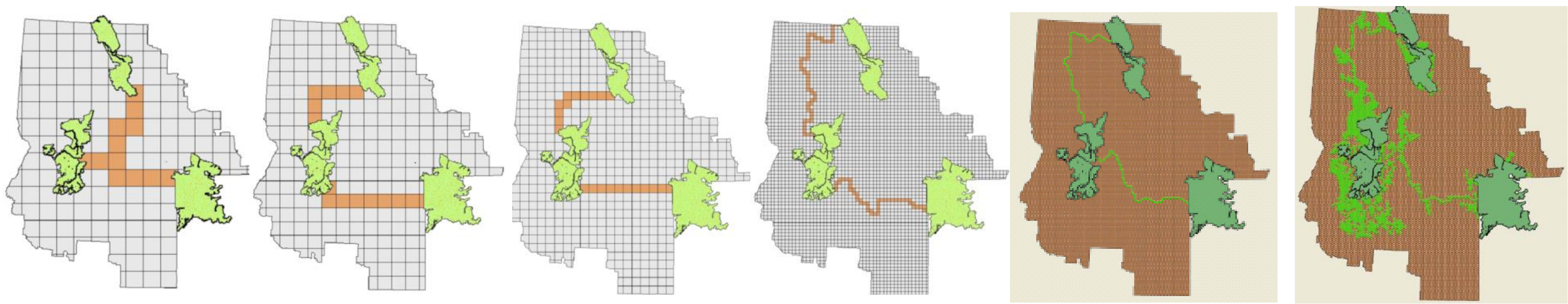
- MIP+Cplex really weak at feasibility testing
- Poor scaling: couldn't even get close to handling real data

Can we do better?

Conrad, G., van Hoeve, Sabharwal, Sutter 2007, 2008

Minimum Cost Corridor for the Connected Sub-Graph Problem

- Ignore utilities → **Min Cost Steiner Tree** Problem
- Fixed parameter tractable – polynomial time solvable for fixed (small) number of terminals or reserves (not used in conservation literature)



50x50 grid	40x40 grid	25x25 grid	10x10 grid	25 km^2 hex	25 km^2 hex
167 Cells	242 Cells	570 Cells	3299 Cells	1288 Cells	Extend with
$1.3B	$891M	$449M	$99M	$7.3M	2xB=$15M
<1 sec	<1 sec	<1 sec	10 mins	2 hrs	10x in Util

What if we were allowed extra budget?

Need to solve problems large number of cells! →Scalability Issues

Understanding Patterns: “Typical” Case Analysis (Synthetic Instances)

How is hardness affected as the budget fraction is varied?

Problem evaluated on semi-structured graphs

m x *m* lattice / grid graph with *k* terminals

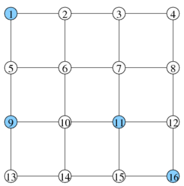
Inspired by the conservation corridors problem

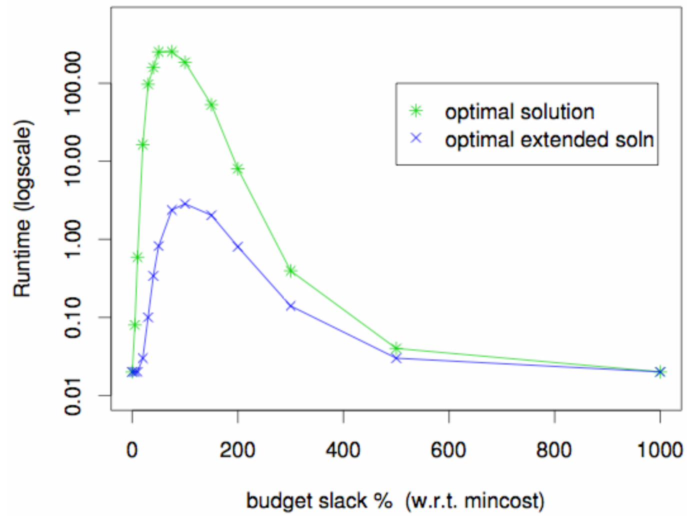
Place a terminal each on top-left and bottom-right

Maximizes grid use

Place remaining terminals randomly

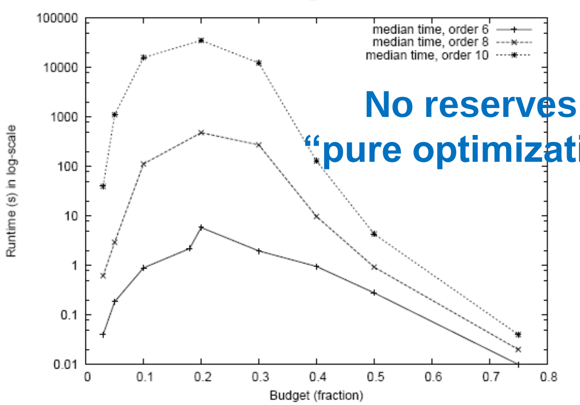
Assign uniform random costs and utilities from {0, 1, …, 10}





3 reserves

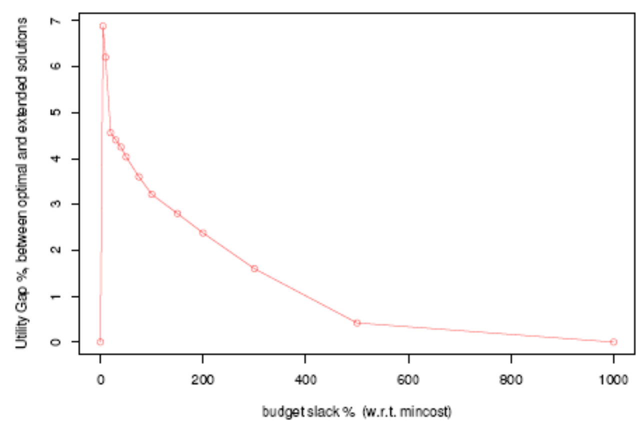
Runtime for Optimal Solution



No reserves: “pure optimization”

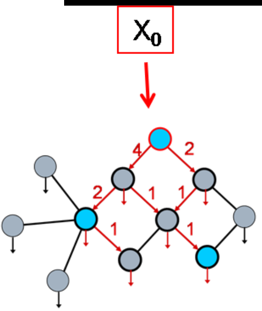
From 6x6 to 10x10 grid (100 parcels): 1000 instances per data-point;

Utility Gap (Optimally Extended Min cost/ Optimal)



3 reserves

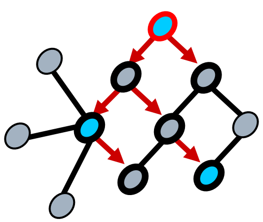
Models Are Important!!!



$$
\begin{aligned}
&\max \sum_{i\in V} u_i x_i \\
&\text{s.t.} \sum_{i\in V} c_i x_i \le C \\
&x_t = 1 \qquad \forall t \in T \\
&x_i \in \{0,1\} \qquad \forall i \in V \\
&\sum_{j\in V-r} x_j = \sum_{i:(r,i)\in E'} y_{ri} \\
&\sum_{i:(i,j)\in E'} y_{ij} = x_j + \sum_{i:(j,i)\in E'} y_{ji} \qquad \forall j \in V \\
&y_{ij} < n x_j \qquad \forall (i,j) \in E' \\
&y_{ij} \ge 0 \qquad \forall (i,j) \in E'
\end{aligned}
$$

Single Commodity Flow

Quite compact (poly size)



$$
\begin{aligned}
&\max \sum_{i\in V} \left(u_i \sum_{j\in\delta(i)} y_{ji} \right) \\
&\text{s.t.} \sum_{i\in V} \left(c_i \sum_{j\in\delta(i)} y_{ji} \right) \le C \\
&\sum_{j\in\delta(i)} y_{ji} = 1 \qquad \forall i \in T, \\
&\sum_{j\in\delta(i)} y_{ji} \le 1 \qquad \forall i \in V - T \\
&y_{ij} + y_{ji} \le 1 \qquad \forall i \in V - T, \forall j \in \delta(i) - r \\
&\sum_{(i,j)\in E' | j\in S, i\in V\setminus S} y_{ij} \ge \sum_{j\in\delta(k)} y_{jk}, \qquad \forall S \subset V - r, \forall k \in S \quad [CUTS] \\
&y_{ij} \in \{0,1\} \qquad \forall (i,j) \in E'
\end{aligned}
$$

Directed Steiner Tree

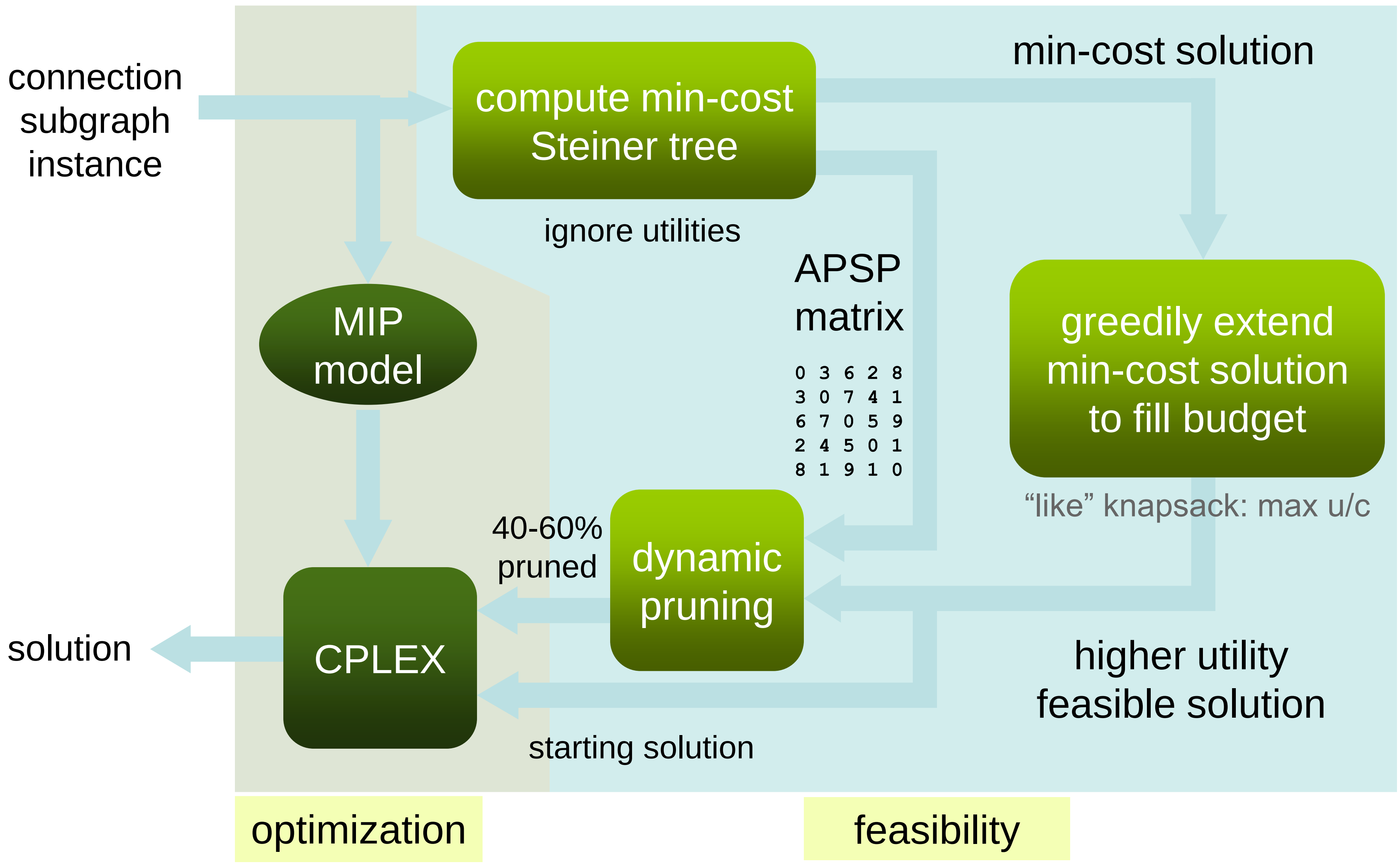
Exponential Number of Constraints ☹!

Captures Better the Connectedness Structure ☺!

Provides good upper bounds
(discover good cuts with ML)☹!

Dilkina and Gomes 2010

Solving the Connection Sub-Graph Problem: Exploiting Structure (A Hybrid MIP/CP Approach)



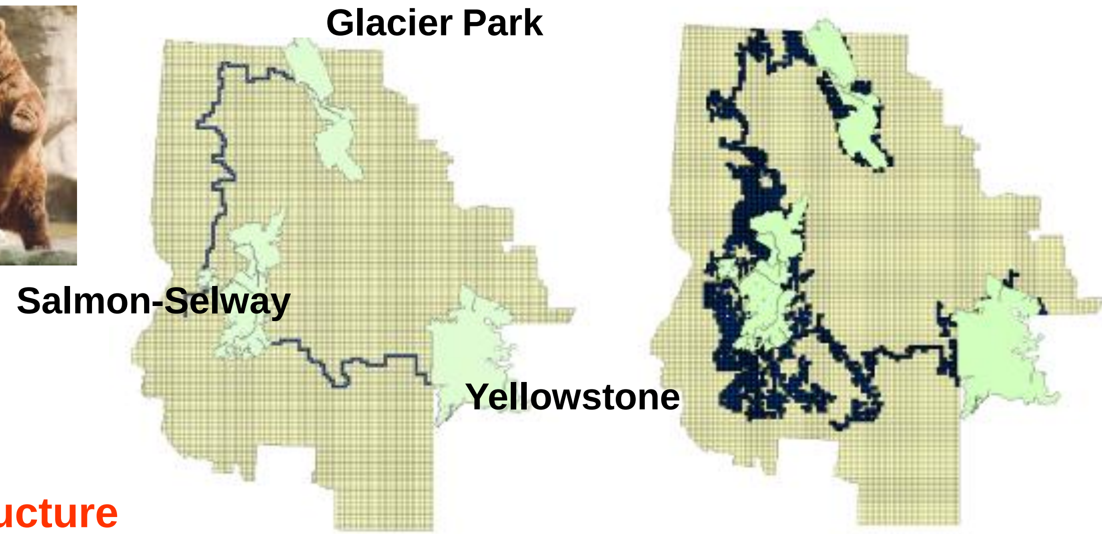
Conrad, G., van Hoeve, Sabharwal, Sutter 2008

Real world instance:

Corridor for grizzly bears in the Northern Rockies, connecting:

Yellowstone
Salmon-Selway Ecosystem
Glacier Park

(12788 nodes)

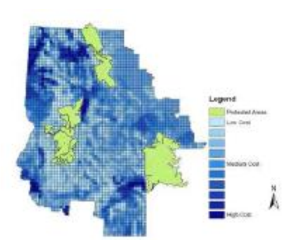


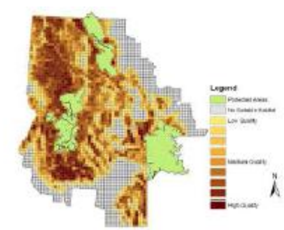
Scaling up Solutions by Exploiting Structure

- Synthetic generator / Typical Case Analysis
- Identification of Tractable Sub-problems
- Exploiting structure via Decomposition
  Static/Dynamic Pruning
- Streamlining for Optimization
- New Encodings

5 km grid
(12788 land parcels):
minimum cost solution

5 km grid
(12788 land parcels):
+1% of min. cost





Our approach allows us to handle large problems and reduce corridor cost dramatically (hundreds of millions of dollars) compared to existing approaches while providing guarantees of optimality in terms of utility:
Optimal or within 1% of optimality for interesting budget levels.

Multiple Species

Wolverines

Other species….

Grizzly Bear

Lynx

Collaborators: Michael K. Schwartz
USDA Forest Service, Rocky Mountain Research Station
Claire Montgomery Oregon State University

Outreach and Education

Boynton Middle School Math Day

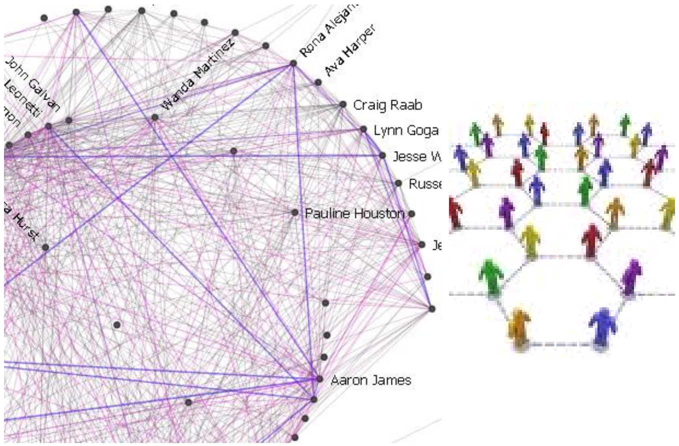
Pedagogical Games
Shortest path, Steiner trees, and much more about Computational Sustainability

"Edutainment" Video Games for middle school

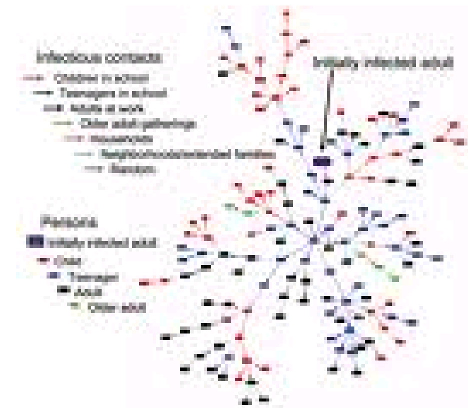
Lots of undergrad/Meng students Having fun designing games for A Travel Museum on Computational Sustainability

Effort led by David Schneider

Connection Subgraph: Other applications



Facebook Network



Network of Pandemic Influenza

Robert J. Glass,* Laura M. Glass,† Walter E. Beyeler,* and H. Jason Min* 2006

- What characterizes the connection between two individuals?
  - The shortest path?
  - Size of the connected component?
  - A "good" connected subgraph?
- Which people have unexpected ties to any members of a list of other individuals?

If a person is infected with a disease who else is likely to be?

[Faloutsos, McCurley, Tompkins '04]

Institute for Computational Sustainability

ICS

Bird Conservation

CORNELL LAB of ORNITHOLOGY

Information Sciences

THE CONSERVATION FUND

America's Partner in Conservation

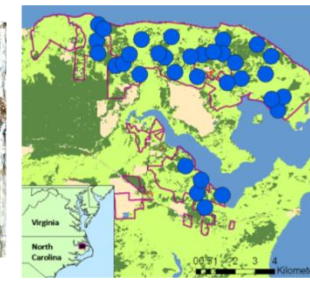
Conservation and Biodiversity: Reserve Design for Bird Conservation

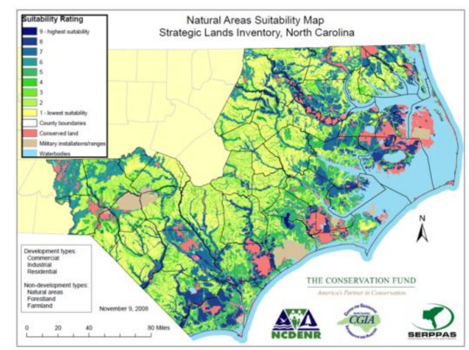
Red Cockaded Woodpecker (RCW) is a federally endangered species

- Current population is estimated to be about 1% of original stable population (~12,000 birds)

Conservation Funds manages

- Palmetto Peartree Preserve (North Carolina)
  - 32 active RCW territories (as of Sept 2008)





Goal: Increase RCW population level

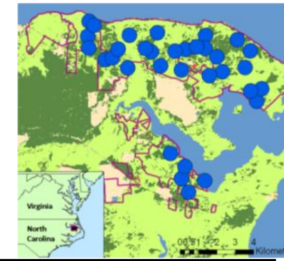
Management options:

- Prioritizing land acquisition adjacent to current RCW populations
- Building artificial cavities
- Translocation of birds

THE CONSERVATION FUND
America's Partner in Conservation

Sheldon, D., Dilkina, B., Ahmadizadeh, K., Elmachtoub, A., Finseth, R., Conrad, J., Gomes, C., Sabharwal, A. Shmoys, D., Amundsen, O., Allen, W., Vaughan, B.; 2009-10

Conservation and Biodiversity: Reserve Design for Bird Conservation



Red Cockaded W.

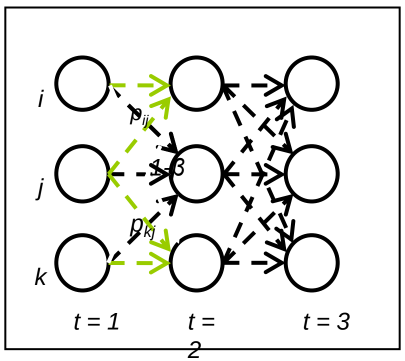
Biological and Ecological Model

Management Decisions:

Land acquisition
Artificial cavities
Translocation of birds

Must explicitly consider interactions between biological/ecological patterns and management decisions

Maximizing the Spread of Cascades



Stochastic diffusion model (movement and survival patterns) in RCW populations

$$\text{maximize } (1/K)\sum_{k=1}^{K}\sum_{i\in\mathcal{R}} x^k(i,T)$$

subject to

$$\begin{aligned}
\sum_{i\in\mathcal{R}}\sum_{t=1}^{T} b(i,t)y(i,t) &\le B; \\
\sum_{t=1}^{T} y(i,t) &\le 1 \quad \forall \text{ territories } i \in \mathcal{R}; \\
x^k(i,t) &\le \sum_{s=1}^{t} y(i,s), \quad \forall \text{ scenarios } k, \text{ territories } i \in \mathcal{R}, \text{ and periods } t; \\
x^k(i,1) &\le a^k(i,i,1), \quad \forall \text{ scenarios } k, \text{ territories } i \in \mathcal{R}; \\
x^k(j,t) &\le \sum_{i\in\mathcal{R}} a^k(i,j,t)x^k(i,t-1), \quad \forall \text{ scenarios } k, \text{ territories } j \in \mathcal{R}, \text{ and periods } 2\ldots T; \\
x^k(i,t) &\ge 0, \quad \forall \text{ scenarios } k, \text{ territories } i \in \mathcal{R}, \text{ and periods } t; \\
y(i,t) &\in \{0,1\}, \quad \forall \text{ territories } i \in \mathcal{R} \text{ and periods } t;
\end{aligned}$$

Stochastic optimization model
Decisions: where and when to acquire land parcels
Goal: Maximize expected number of surviving RCW

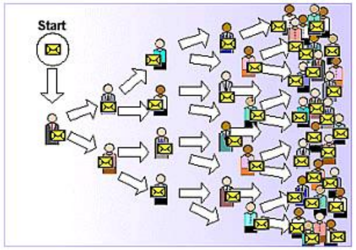
Computational Challenge: scaling up solutions for considering a large number of years (100+)
→ decomposition methods and exploiting structure

Maximizing (Minimizing) Cascade Spread
Other applications

- A lot of phenomena can be *modeled* as a stochastic diffusion process on a network of nodes

  – Social Networks:
    - Technology adoption/ Viral marketing
    - Rumors / News / Gossip
    - Disease

  – Invasive Species



Viral Marketing

- We often can and want to intervene to influence the expected outcome of the diffusion process - *optimization*

[Domingos and Richardson, KDD 2001]

[Kempe, Kleinberg, and Tardos, KDD 2003]

[Krause and Guestrin 2007]

Maximizing Spread of Cascades

- General solution methodology – Sample Average Approximation --- for optimizing management decisions to maximize the expected spread of a cascade
  - SAA - Reduces stochastic optimization problems to solving a deterministic analogue
  - Samples a finite set of scenarios from the underlying distribution
  - Provides guarantees
- Approach outperforms previous approaches based on greedy methods (exploiting submodularity which does not apply to most cascade problems)

Maximizing
the Spread of Cascades @ UAI

Sheldon, D., Dilkina, B., Ahmadizadeh, K., Elmachtoub, A., Finseth, R., Conrad, J., Gomes, C., Sabharwal, A. Shmoys, D., Amundsen, O., Allen, W., Vaughan, B.; 2009-10

Additional Levels of Complexity: Stochasticity, Uncertainty, Large-Scale Data Modeling

- How to estimate population distributions and habitat suitability?

  - Multiple species (hundreds or thousands), with interactions (e.g. predator/prey).

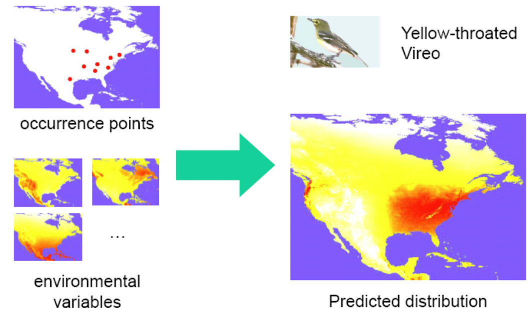
  - Movements and migrations;

  - Climate change

  - How to get the data?

- Other factors

(e.g., different models of land conservation (e.g., purchase, conservation easements, auctions) typically over different time periods).



Maxent

Steven Philips, Miro Dudik & Rob Schapire

Challenging Problems
at the **Intersection of**
Machine Learning and Optimization

Citizen Science: EBird

CORNELL LAB of ORNITHOLOGY

Information Sciences

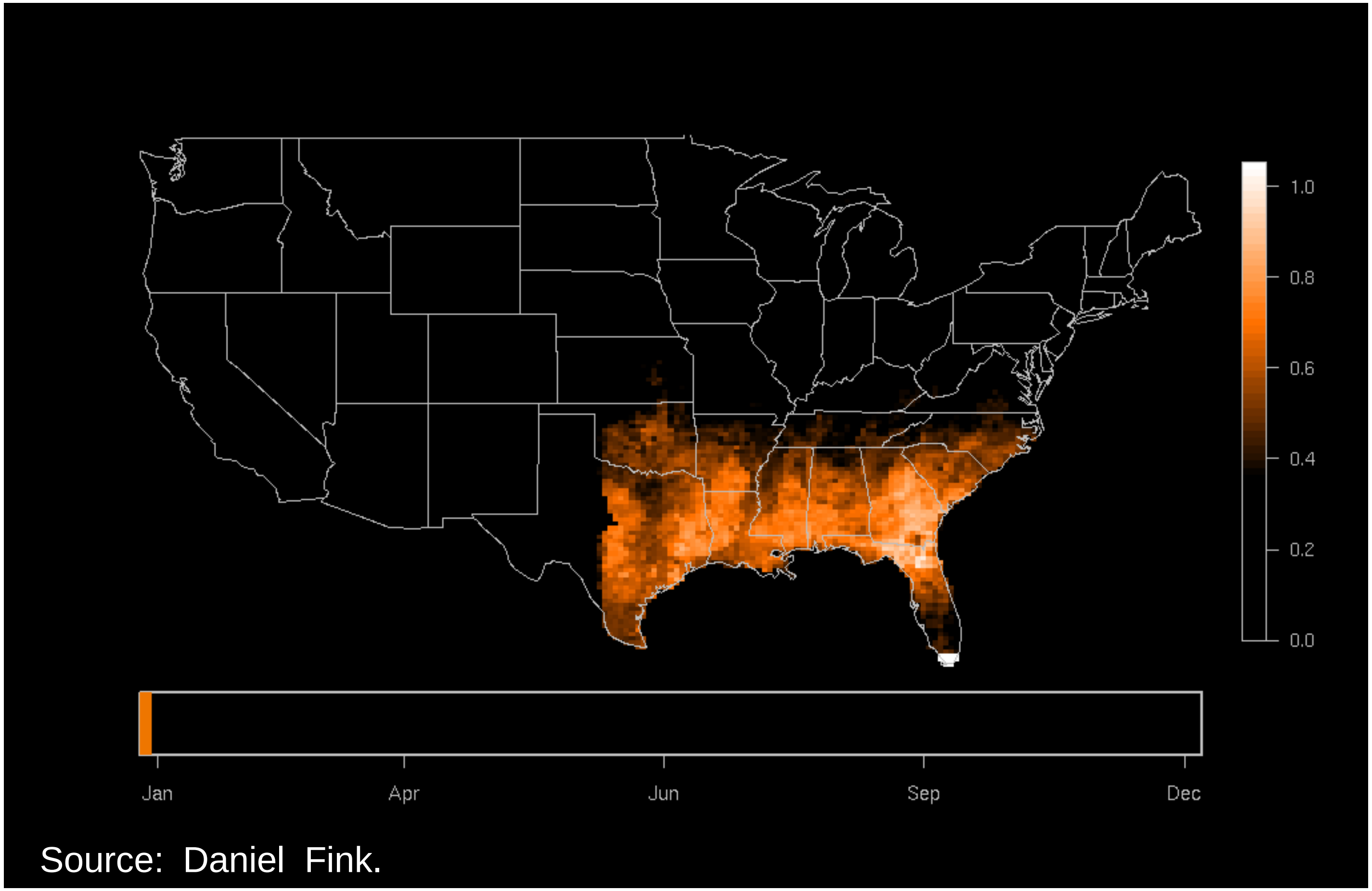
OSU

Tom Dietterich, Jo Albers, and Weng-Keen

Ebird: Citizen Science @ Lab. Of Ornithology @ Cornell

http://www.birds.cornell.edu/

Eastern Phoebe Migration



Information Sciences

Daniel Fink,Wesley Hochachka, Art Munson, Mirek Riedewald, Ben Shaby, Giles Hooker, and Steve Kelling, 2009.

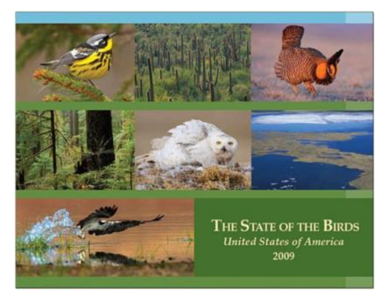
Wind Energy and Bird Conservation

Existing and proposed wind farms in US and Mexico (2008)

- 26,000+ turbines, 1.5% of potential “Build-out” to reach potential would require >1.7 million turbines
- Areas with most favorable winds are also often associated with migratory pathways

Where to not locate wind farms?

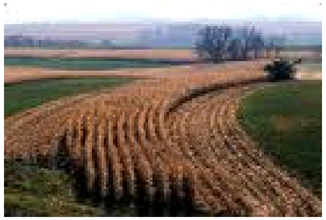
Guidelines for locating wind farms



Ken Rosenberg (ICS) part of the science team for the State of the Birds Report

Andrew Farnsworth, Ken Rosenberg, 2009

Socio-Economic Systems:
Poverty, Agricultural Systems, Harvesting Policies



Poverty maps

Identifying the poor is the first essential step

<u>Poverty maps:</u>

- Use multiple data sets to estimate and map poverty patterns not directly measured.
- Machine learning and related methods can permit more efficient use of data from varied sources.

New collaboration economists and computer scientists and social scientists @ ICS

Example: 2002 Uganda poverty map

Chris Barrett and Corey Lang

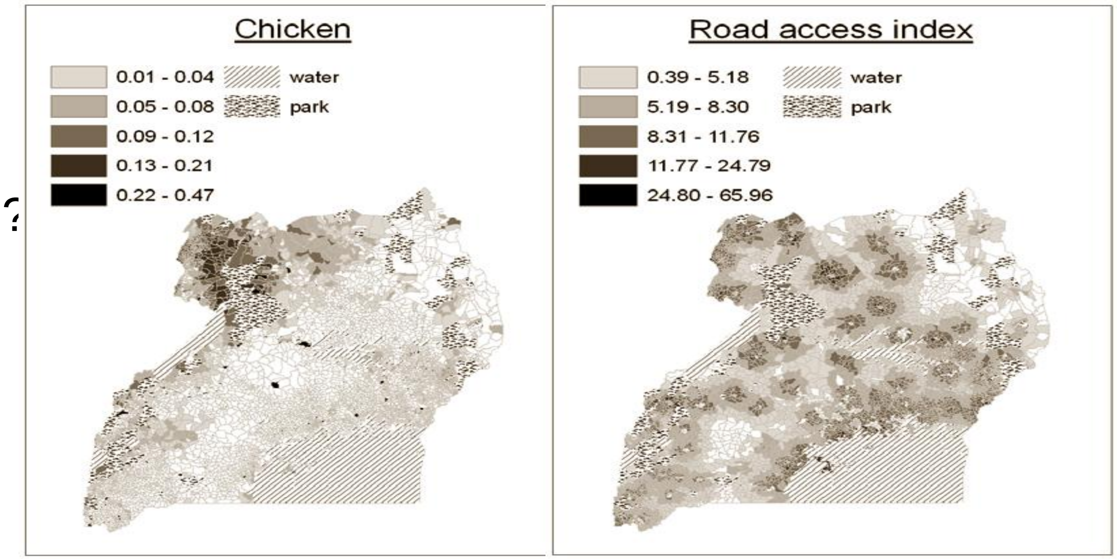
Targeting maps

Targeting the best response to reduce poverty

Targeting maps:

How to estimate the impact and marginal returns of different assets?

- Poverty interventions need to be targeted to specific areas
- Asset-based investments have spatially-varying marginal returns



Average marginal returns of different assets (Uganda)

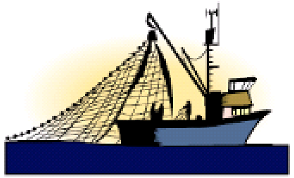
Policies for Poverty Reduction:
Which set of interventions to apply to each targeted area
to maximize the poverty reduction, subject to resource constraints (e.g. budget))?

Challenging problems at the intersection of machine learning and optimization, applied economics and social science.

Chris Barrett and Corey Lan

Harvesting Policies
Fishery Management

- The biomass of fish is estimated to be 1/10 of what it was 50 years ago and is declining (Worm et al. 2006).



- The state of the world's marine fisheries is alarming
- Researchers believe that the collapse of the world's major fisheries is primarily the result of the mismanagement of fisheries (Clark 2006; Costello et al. 2008).

There is therefore a clear urgency of finding ways of
defining policies for managing fisheries is a sustainable manner.

Natural Resource Management: Policies for harvesting renewable resources

Economy

Y_t

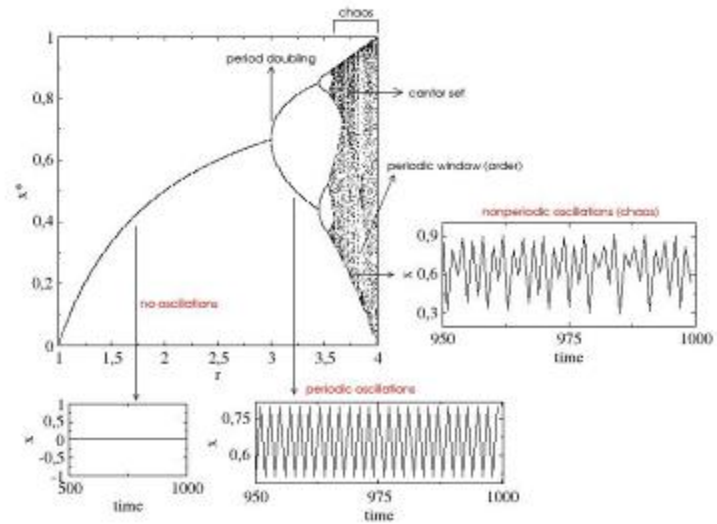
Harvest of a Renewable Resource: Tuna

$$X_{t+1} - X_t = F(X_t) - Y_t$$

X_t = the fish stock (tuna)
Y_t = the rate of harvest
$F(X_t)$ = the net growth function

Non-linear dynamics

Example of a Biological Growth Function F(x):
Logistic map: $x_{t+1} = r\, x_t (1 - x_t)$, r is the growth rate



Increasing Complexity: more complex models and multiple species interactions

We are interested in identifying policy decisions
(e.g. when to open/close a fishery ground over time).

Uncharted territory: Combinatorial optimization problems with an underlying dynamical model.

Class of Computationally Hard Hybrid Dynamic Optimization Models

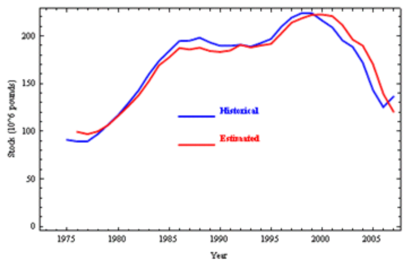
Clark 1976; Conrad 1999; Ang, Conrad, Just, 2009; Ermon , Conrad, Gomes and Selman 2010

An Application to the Pacific Halibut Fishery in Area 3A
Regulated by International Pacific Halibut Commission (IPHC)

Population Dynamics

- Beverton-Holt Model

$$x_{n+1} = f(s_n, w_n) = (1-m)s_n + w_n \frac{r_0 s_n}{1 + s_n/M}.$$



x_n	stock
h_n	harvest
$s_n = x_n - h_n$	escapement
w_n	stochasticity
m	mortality
r_0	proliferation rate
M	capacity

Economics Model

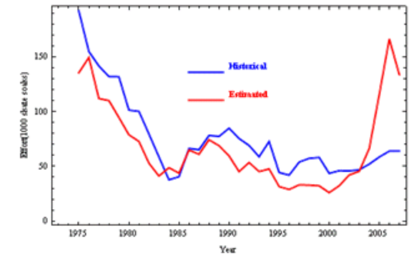
- Revenue $ph - \int_{x-h}^{x} g(y)dy - K$

h	harvest
p	price
	fixed costs
g(y)	marginal cost
q , b	elasticities

- Variable costs proportional to **effort** E_t ("skate soaks")

- Model for the **effort**:

$$E_t = \int_{x_t - h_t}^{x_t} \frac{1}{qy^b} dy$$



Optimization problem

Mini-max formulation

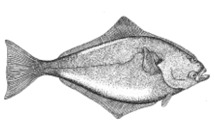
$$C_N^{\pi}(x) = \min_{\substack{w_1,\dots,w_N \\ w_i \in W(x_i)}} \sum_{n=1}^{N} \alpha^n \left(R(x_n) - R(x_n - h_n) - K\delta_0(h_n)\right)$$

1 step revenue

Discount rate

s.t.

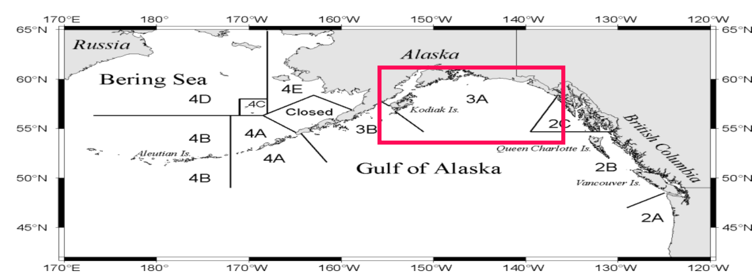
$$x_{n+1} = f(x_n - h_n, w_n) \qquad h_n = \mu_n(x_n)$$



→Admissible Policies $\pi = \{\mu_1, \dots, \mu_N\}$ $0 \le \mu_i(x) \le x$

→Robust (Risk averse) formulation

→No fine grained characterization of randomness needed (**Game against nature**)



Pacific Halibut Fishery in Area 3A

Ermon , Conrad, Gomes and Selman UAI 2010

General MDP Framework

Resource dynamics

$$x_{n+1} = f(x_n - h_n, w_n)$$

New stock size — Amount left — Stochasticity

Resource growing over time
Uncertain dynamics
Economics of harvesting

Closed loop policy

$$\pi = \{\mu_i\}_{i \in [1,N]} \quad h_n = \mu_n(x_n)$$

To maximize total discounted utility

$$\sum_{t=0}^{N} \pi(x_t,\ h_t)\alpha^t$$

Goal: policy for optimal exploitation over time

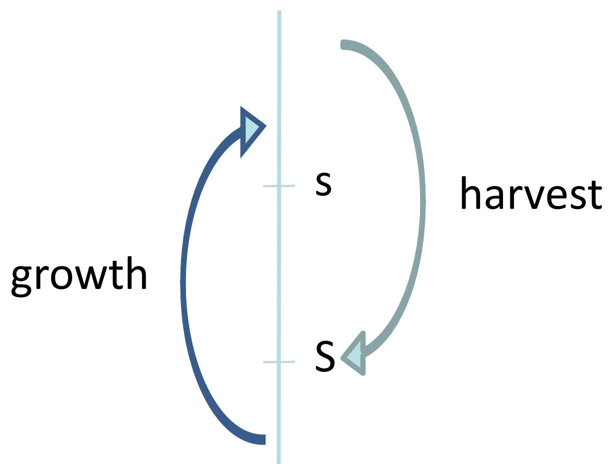
→ natural **MDP formulation**

Ermon , Conrad, Gomes and Selman UAI 2010

Main result

- **Fixed costs** and **variable costs**
- **Concavity** of the growth function
- **Robust** optimization framework (Worst-case for Nature)

Theorem: The optimal policy is of S-s type.



- If the stock is greater than s, harvest down to S
- If the stock is smaller than s, let the stock grow

The Pacific Halibut Fishery in Area 3A

- Regulated by International Pacific Halibut Commission (IPHC)

Total Allowable Catch Policy →
Constant Proportion Policy
→rate of fishing ~ 12.3% of **estimated stock**.

Our Result: Policies with periodic closures of a fishery can outperform Constant Proportion Policies

Ermon, Conrad, Gomes, Selman, UAI 2010 **Optimal policies for management of renewable resources using MDPs**

Markov Decision Processes (MDPs) to Model for many **renewable resource allocation problems.**

General Approach to model renewable resources with complex dynamics!!!!


Renewable resources:

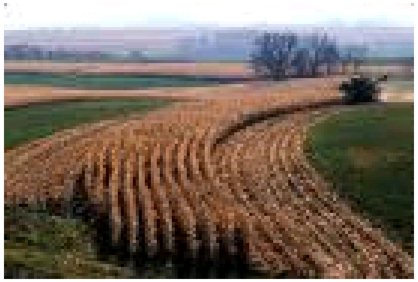
forests and **fisheries**

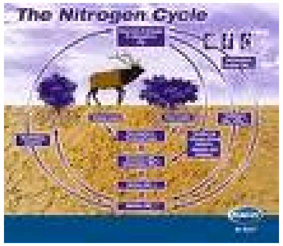
Problems with a similar mathematical structure:

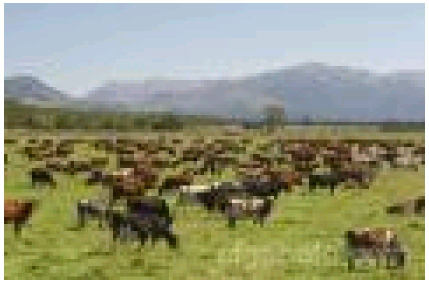
- **Pollution management,**
- **Invasive species control,**
- **Supply chain management** and
- **Inventory control, and many more.**


Ermon, Conrad, Gomes, Selman, UAI 2010 **Optimal policies for management of renewable resources using MDPs**

Agricultural Sustainability and Sustainable Communities

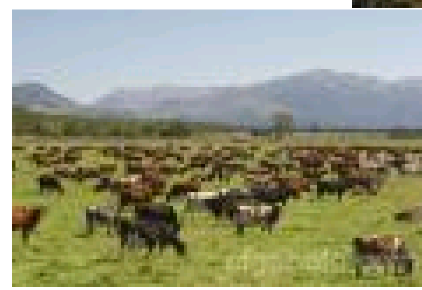


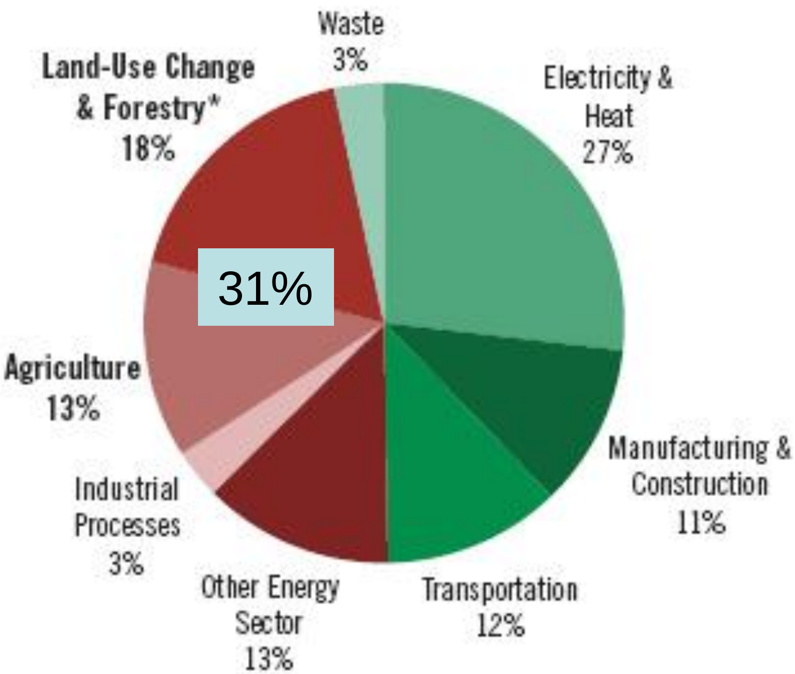




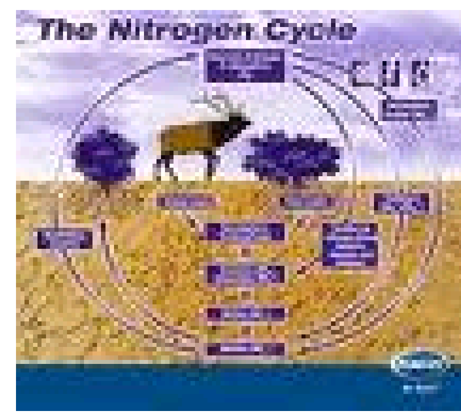
Agricultural Systems and GHG's emissions

Agriculture is the ***primary driver of land use change and deforestation*** *and an important* ***source of*** **Greenhouse Gases** (GHG)s: 52% and 84% of global anthropogenic CH_4 and N_2O, significant $C0_2$



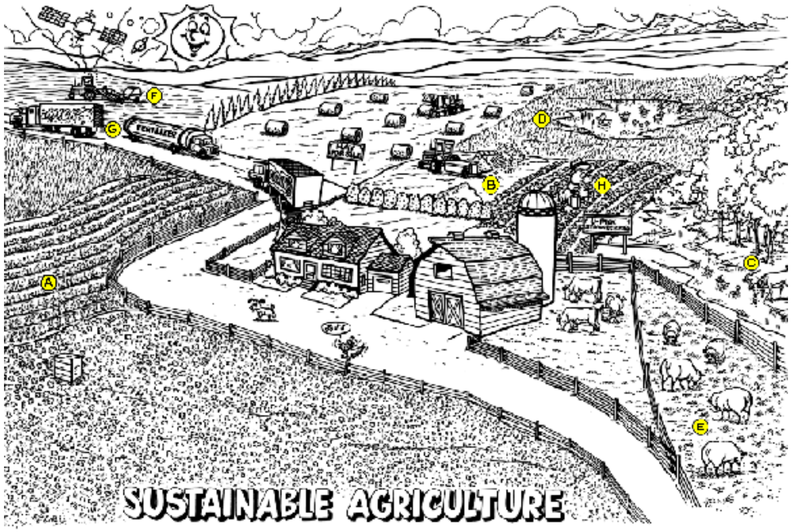


Nitrogen Based Fertilizers

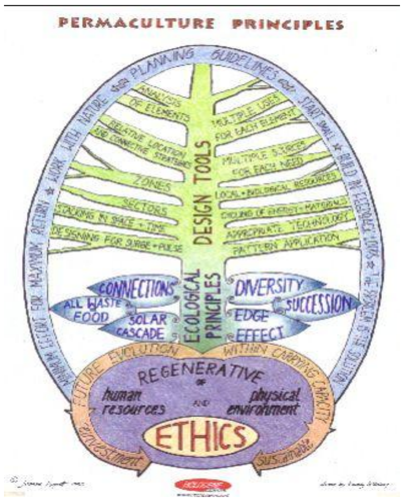


Source of Greenhouse Gases

Sustainable Agriculture and Sustainable Communities



Super eco



Super Techno

Sustainable Agriculture and
Sustainable Communities
Collaborators
Laurie Drinkwater and Norm Scott

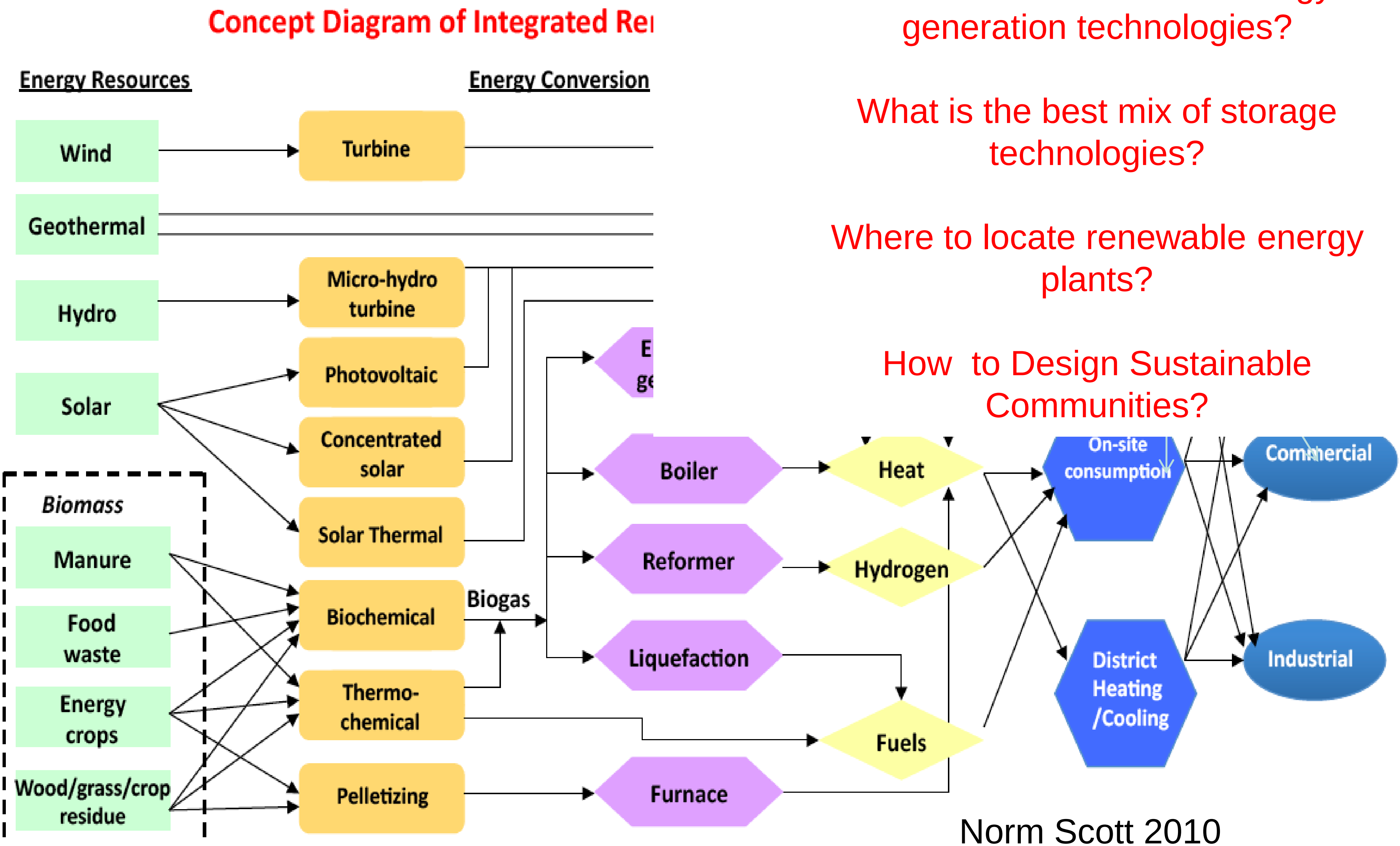
Renewable Energy Systems for Sustainable Communities

What is the best mix of energy generation technologies?

What is the best mix of storage technologies?

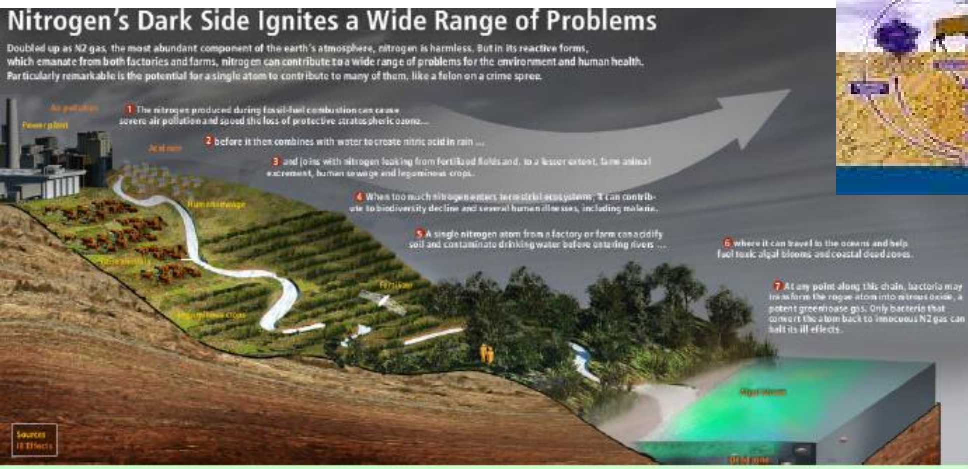
Where to locate renewable energy plants?

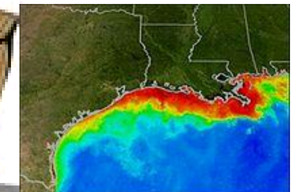
How to Design Sustainable Communities?



Norm Scott 2010

Nitrogen Cycle and Fertilizers





Nitrogen Based Fertilizers

Dead Zone in Gulf of Mexico

Townsend & Howarth (2010) Scientific American

Fertilizers have played a key role in the increase of food production but they key culprits in the production of greenhouse gases emissions creating dead zones

Spatial and temporal analysis of

Nitrogen cycle

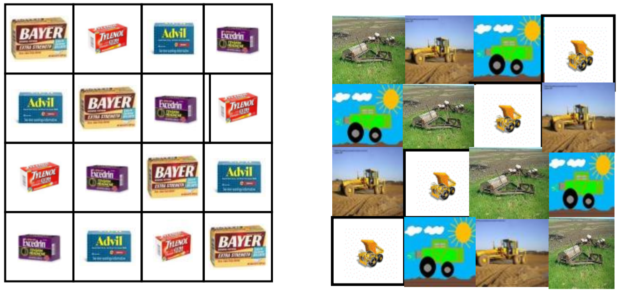
(Collaborator Bob Howarth)

Study of fertilizers and

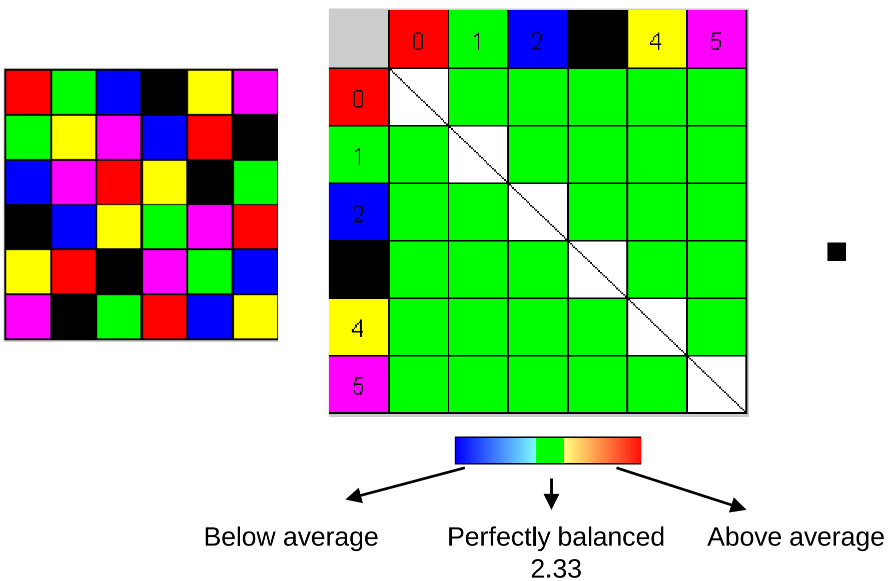
design of Experiments

Collaborator Harold van Es

Design of Scientific Experiments (4 Treatments: A,B,C,D)



Latin Square



Spatially Balanced Latin Square

- Hybrid IP/CSP based
  - Assignment formulation
  - Packing formulation
  - Different CSP models
- SAT/ CSP based approach
  - State of the art model for Latin Square

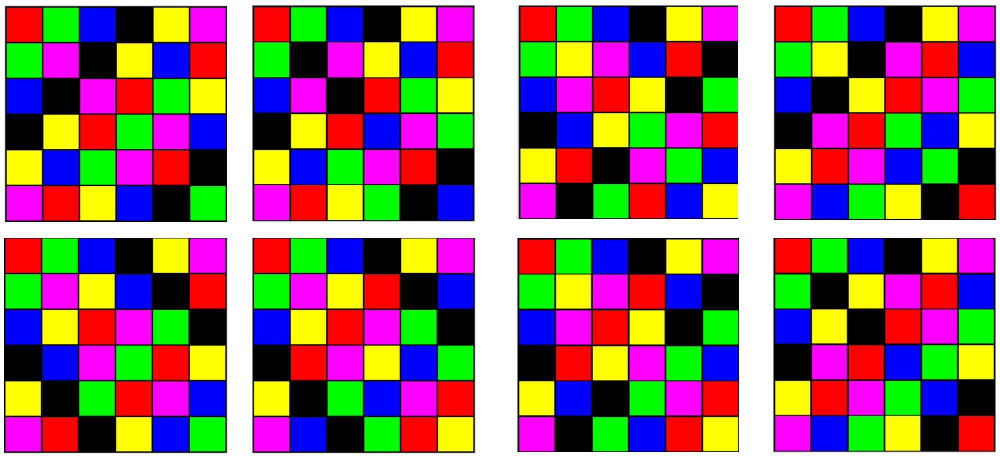
  + symmetry breaking by initializing first row and column (SBDD doesn't help; this is not a completion problem)
- Local search based approach

These approaches do not scale up

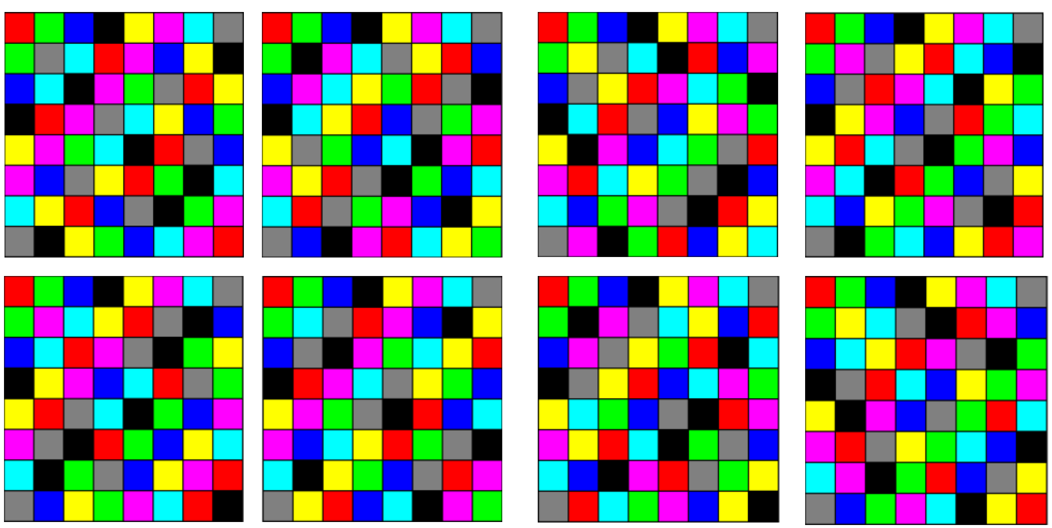
max order 6

(Target number 30).

Diagonally Symmetric SBLS



Observed symmetric SBLS of Order 6

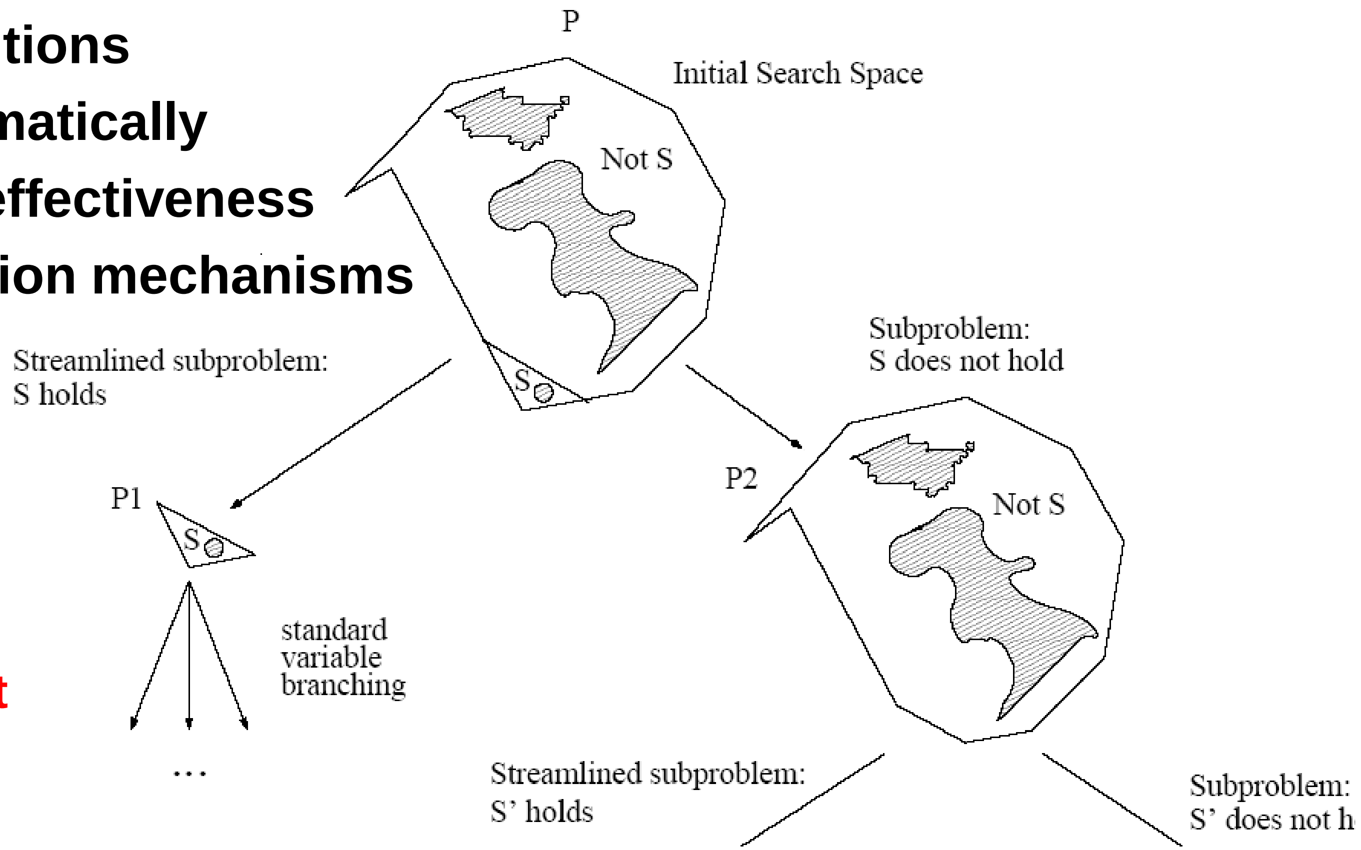


Imposing symmetry

could generate
SBLS of order 8 and 9

Streamlining in Terms of Global Search

Goal: Exploit the structure of solutions to dramatically boost the effectiveness of the propagation mechanisms



P1 Substantially smaller than its complement P2

Streamlining: strong branching mechanisms at high levels of the search tree.

Science of Computation: Discovering patterns, laws, and hidden structure in computational phenomena

Streamlining Constraint Reasoning

→Streamlining
→Scaling up of solutions

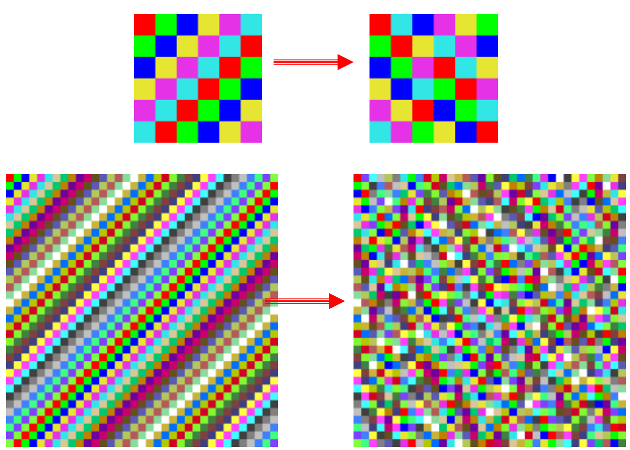
Discovery of structural properties across solutions (machine learning);

Divide (“streamline”) the search space by imposing such additional properties.

Design of Scientific Experiments for Studying Fertilizers – Spatially Balanced Latin Squares

Existence of SBLS – open problem in combinatorics

Streamlining by Permutation of Complete Columns of a Cyclic Latin Square



SBLS Order 35

Gomes and Sellmann 2003/2004

Discovery of Construction for SBLS for experimental design

LeBras, Perrault, Gomes, 2010

Can we find an domain independent way of streamlining?

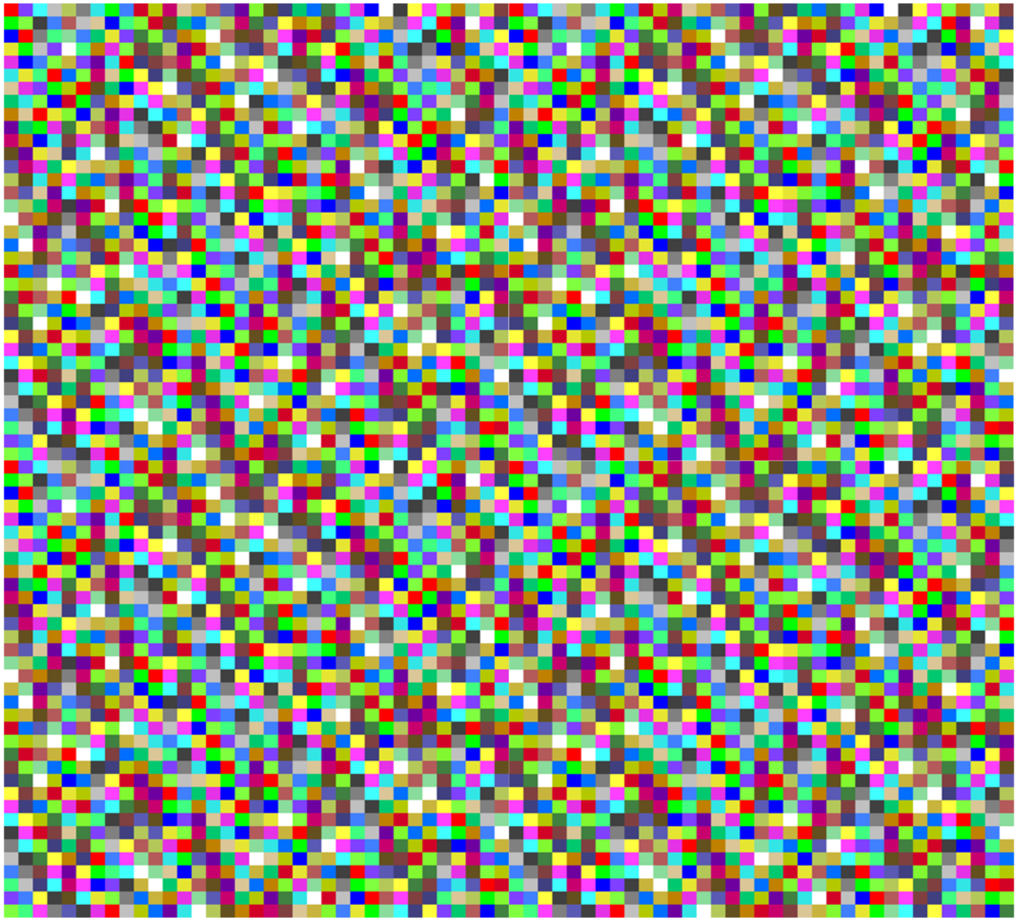
YES: XOR-Streamlining based on random parity constraints

(Valiant and Vazirani 1986, Unique SAT)

Provable bounds for Counting and Sampling.

Gomes, Sabharwal, Selman, 2006

Discovery of Construction for SBLS for experimental design



SBLS Order **932**

LeBras, Perrault, Gomes, 2010

Inter-disciplinary Research Project:
Material Discovery for Fuel Cell Technology

Material Discovery for Fuel Cell Technology

Goal of Material Discovery:

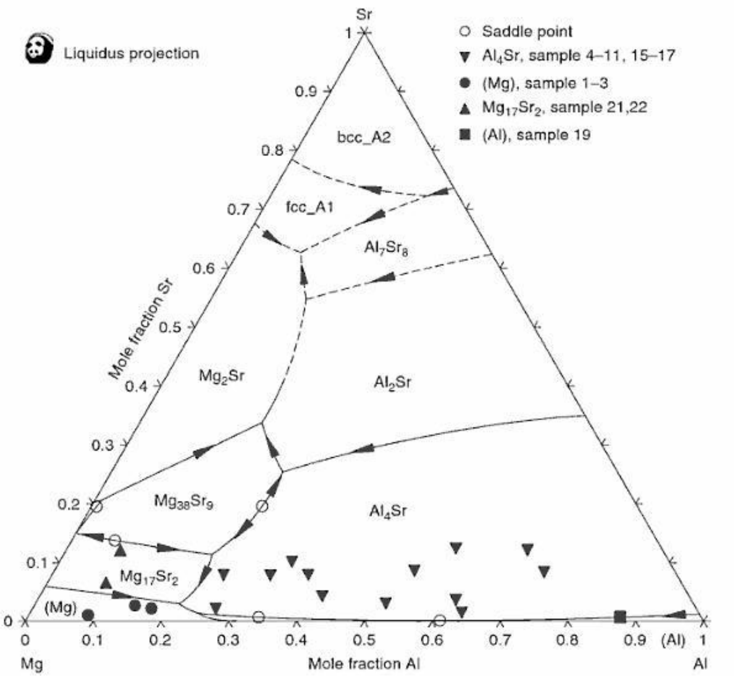
- Find new products
- Find product substitutes
- Understand material properties

Approach: Analysis of inorganic libraries

→Sputter 2 or 3 metals (or oxides) onto a silicon wafer (which produces a thin-film)

→Use x-ray diffraction to study crystallographic structure of the thin-film

→Note: electromagnetic radiation experiments are very expensive!!

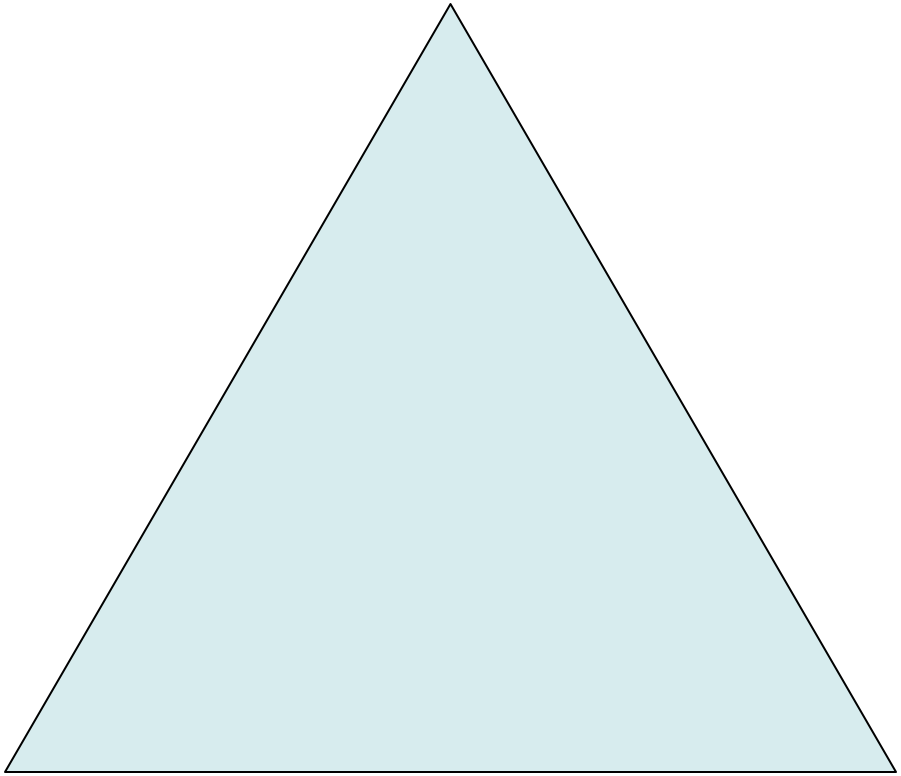


Ronan LeBras, Damoulas, John M. Gregoire, Sabharwal, Gomes, van Dover

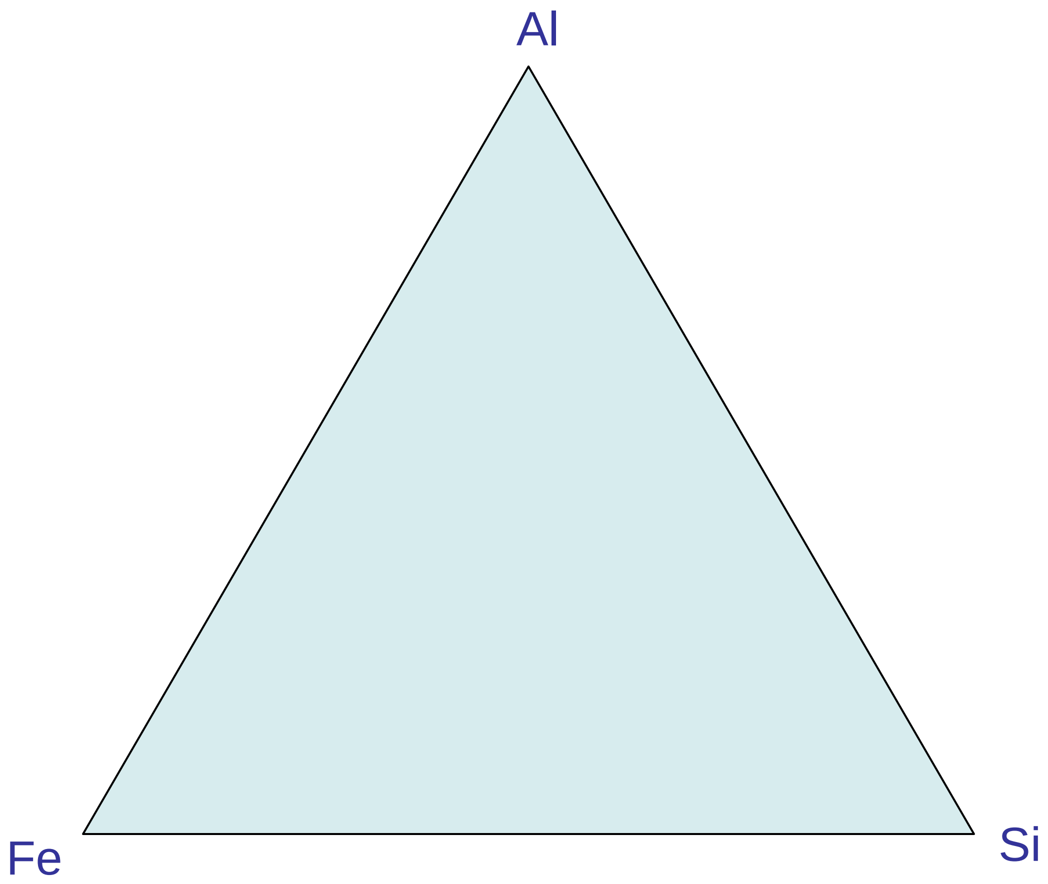
Example: study of platinum-tantulum library showed correlation between crystallographic phase and improved fuel cell oxidation catalysis (Gregoire et al 2010)

Fuel Cell Applications

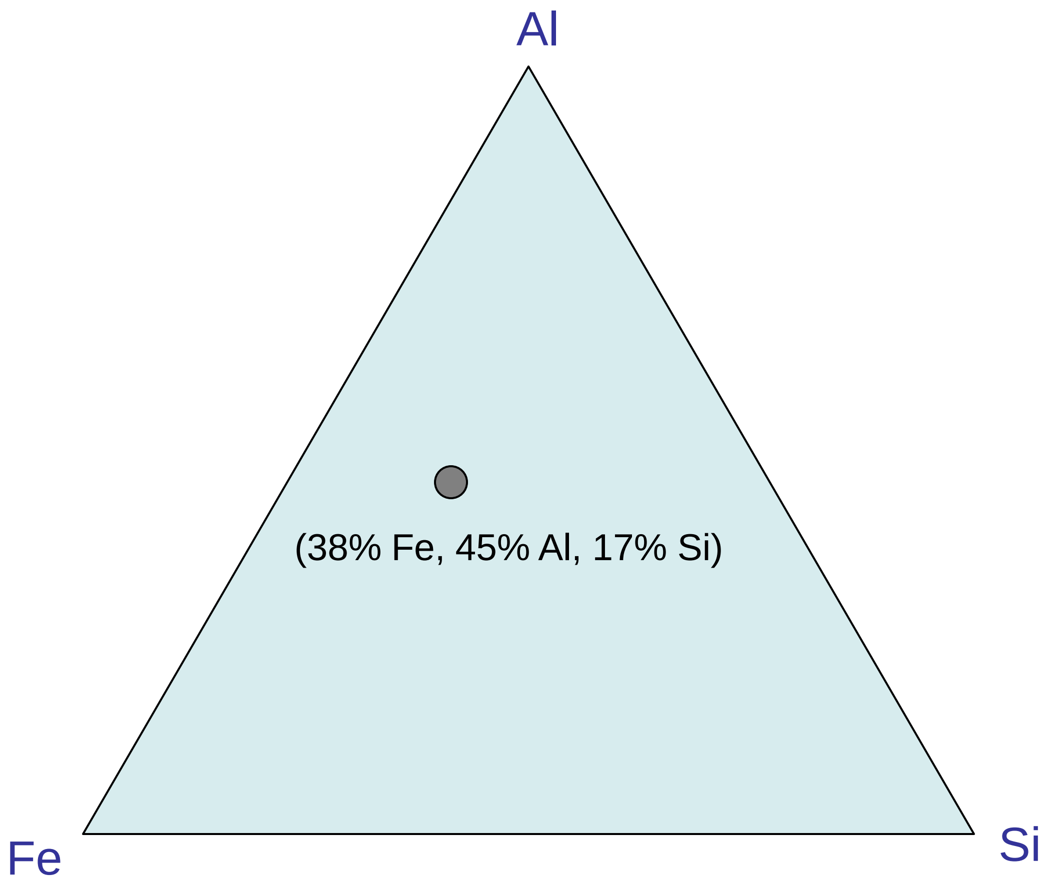
Problem Definition



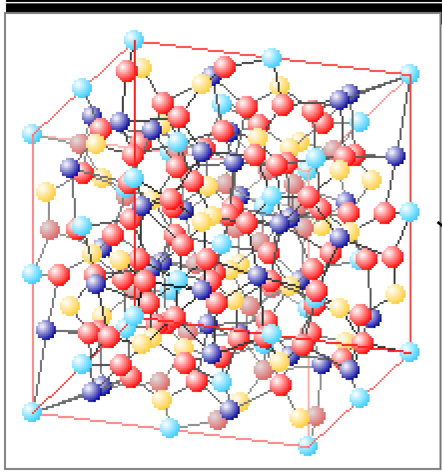
Problem Definition



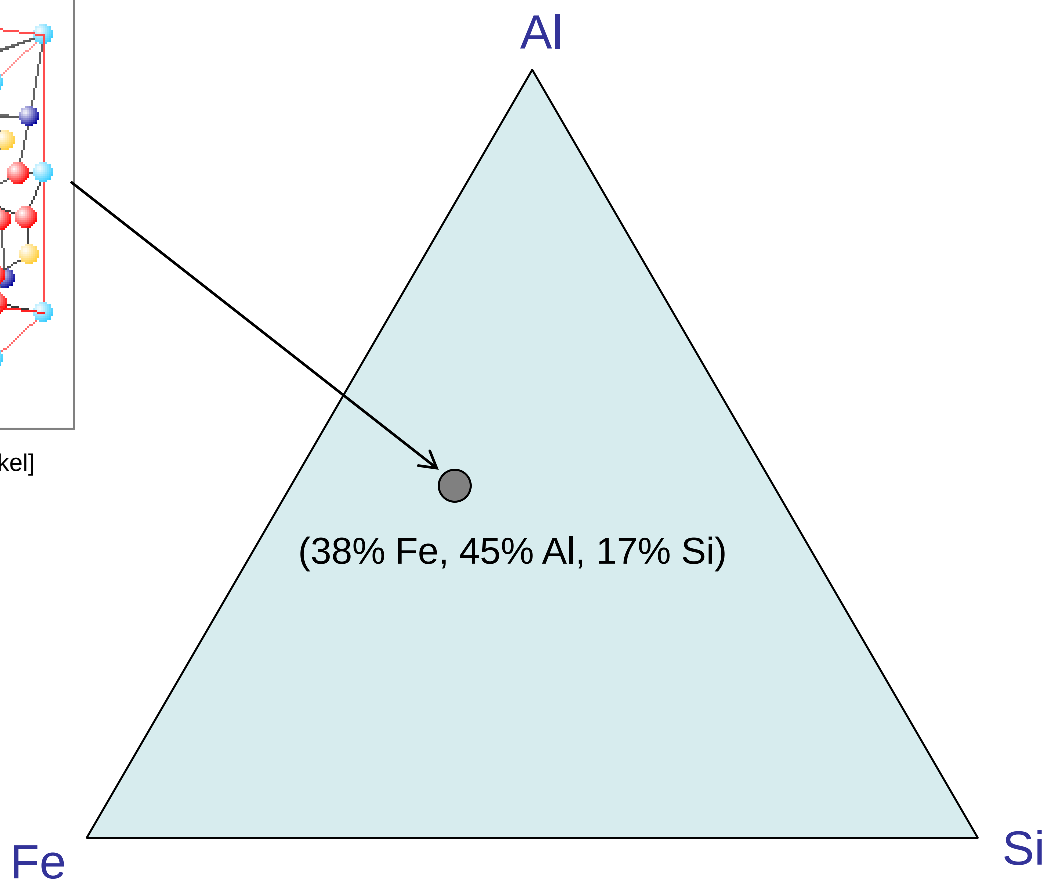
Problem Definition



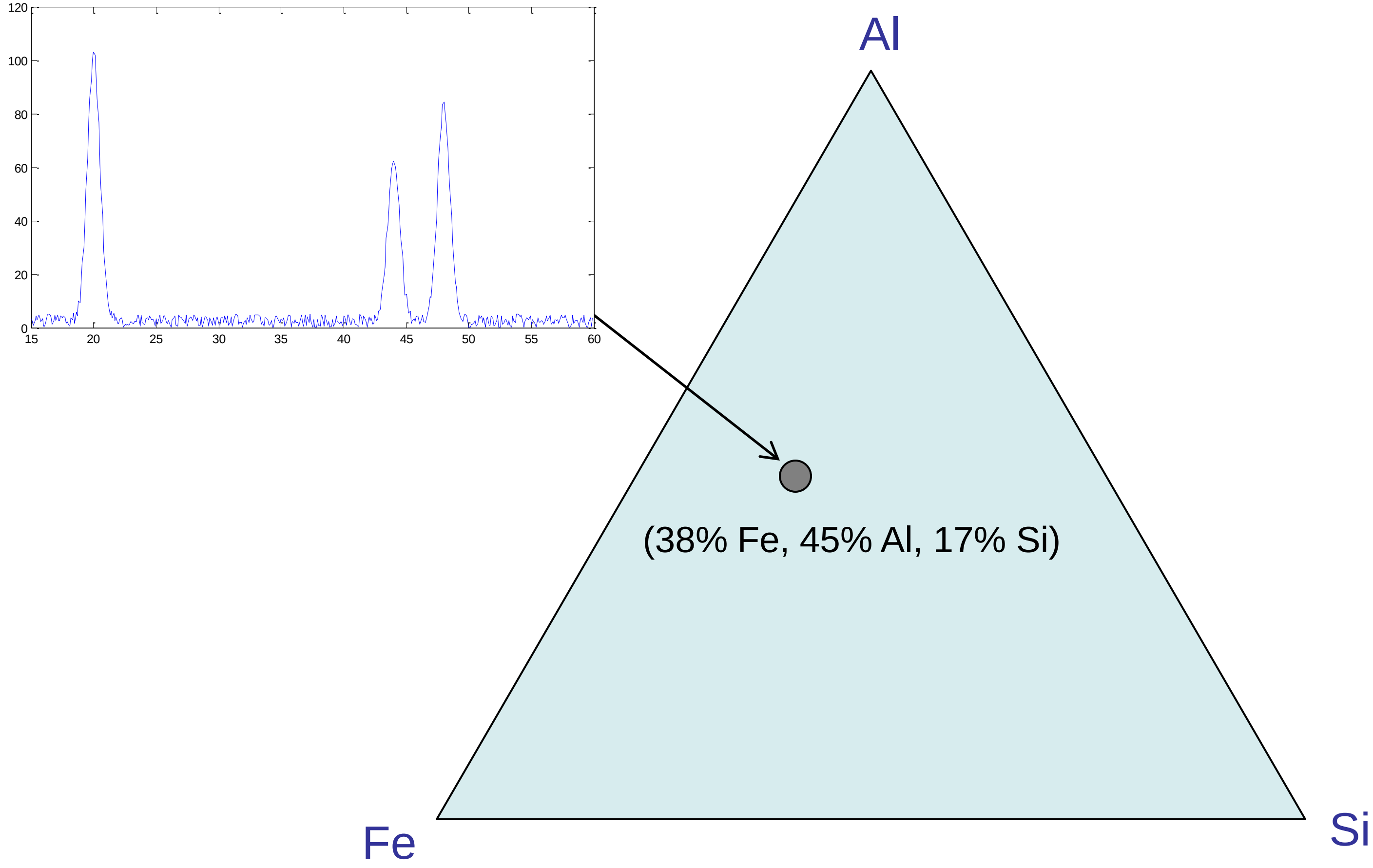
Problem Definition



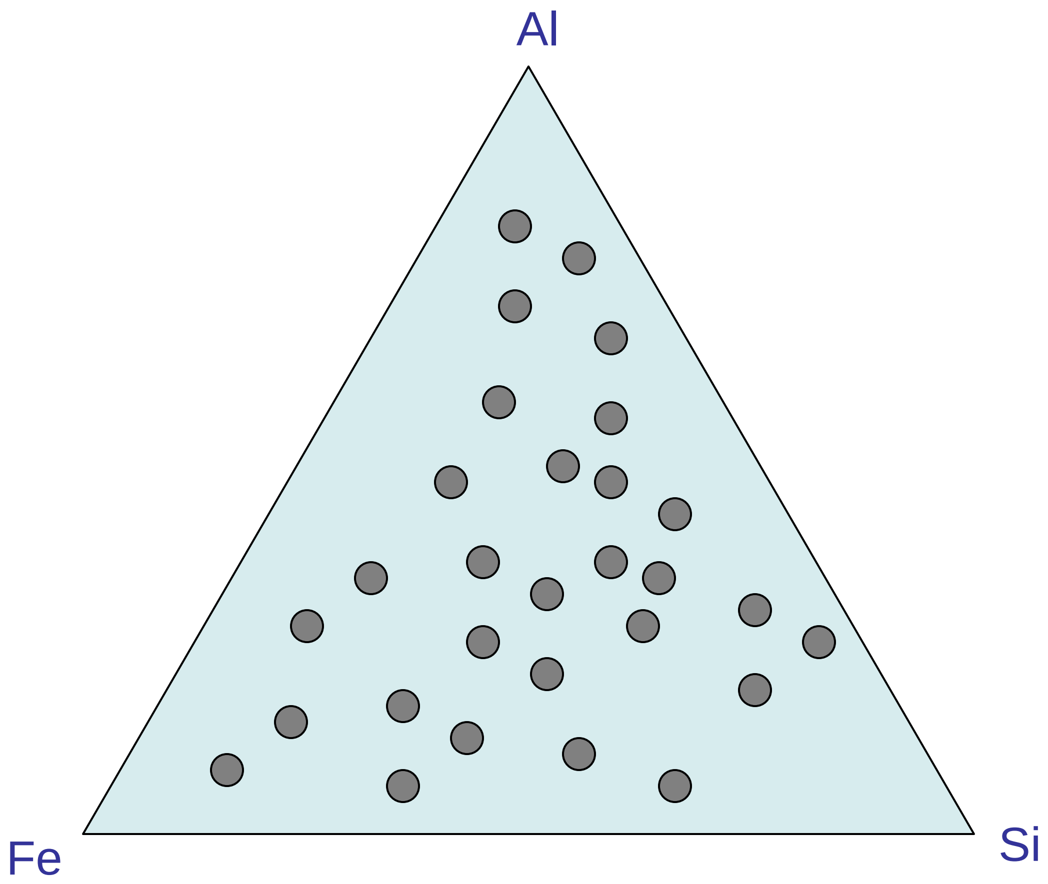
[Source: *Pyrotope*, Sebastien Merkel]



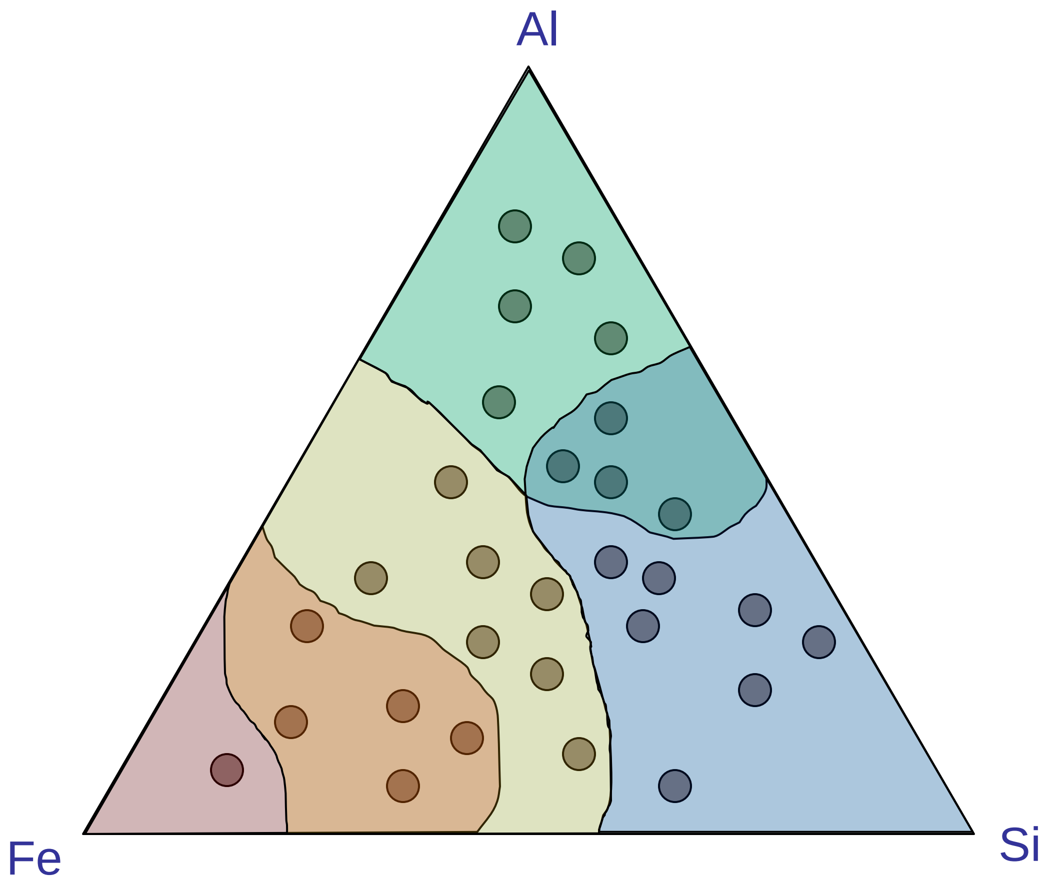
Problem Definition



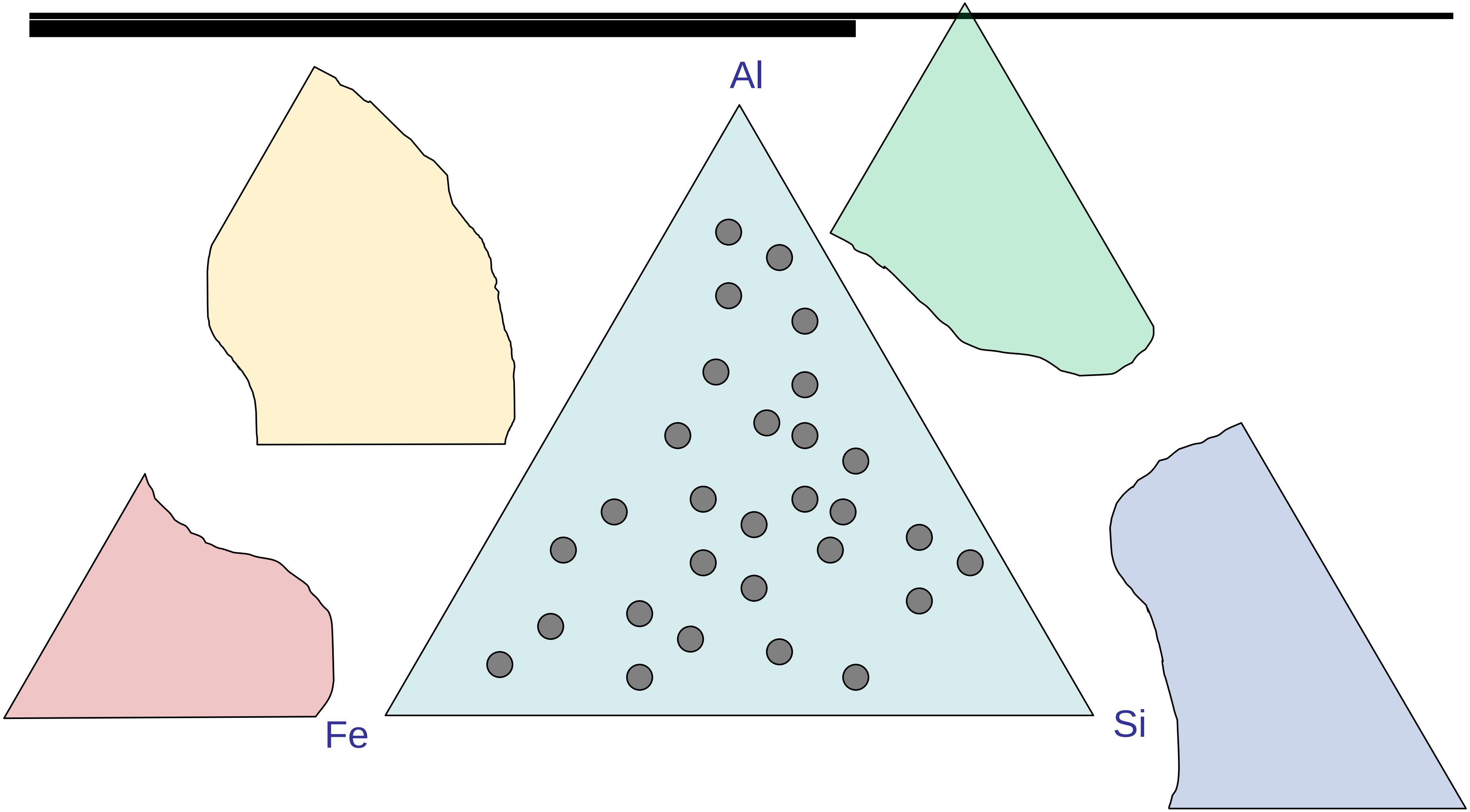
Problem Definition



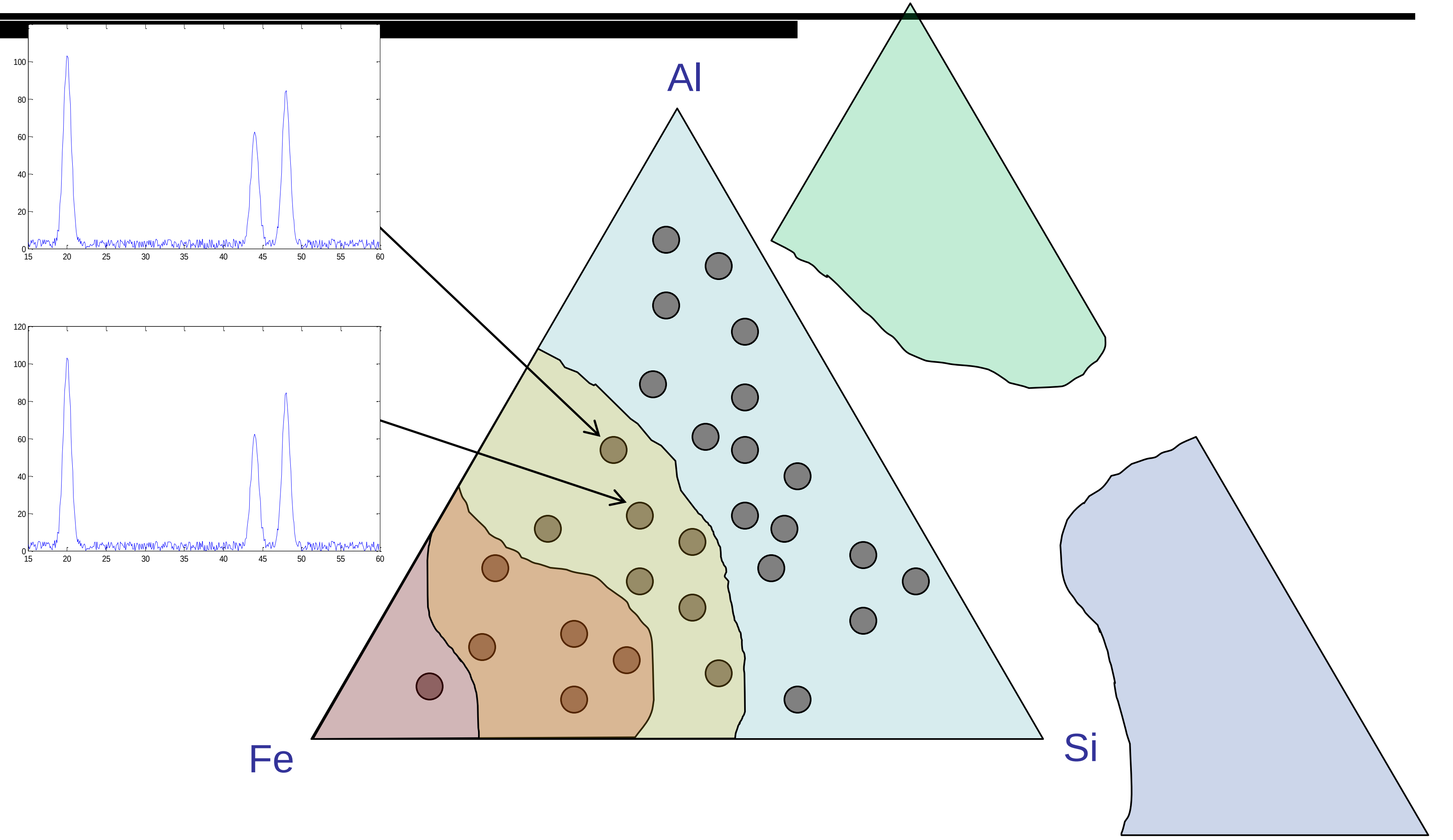
Problem Definition



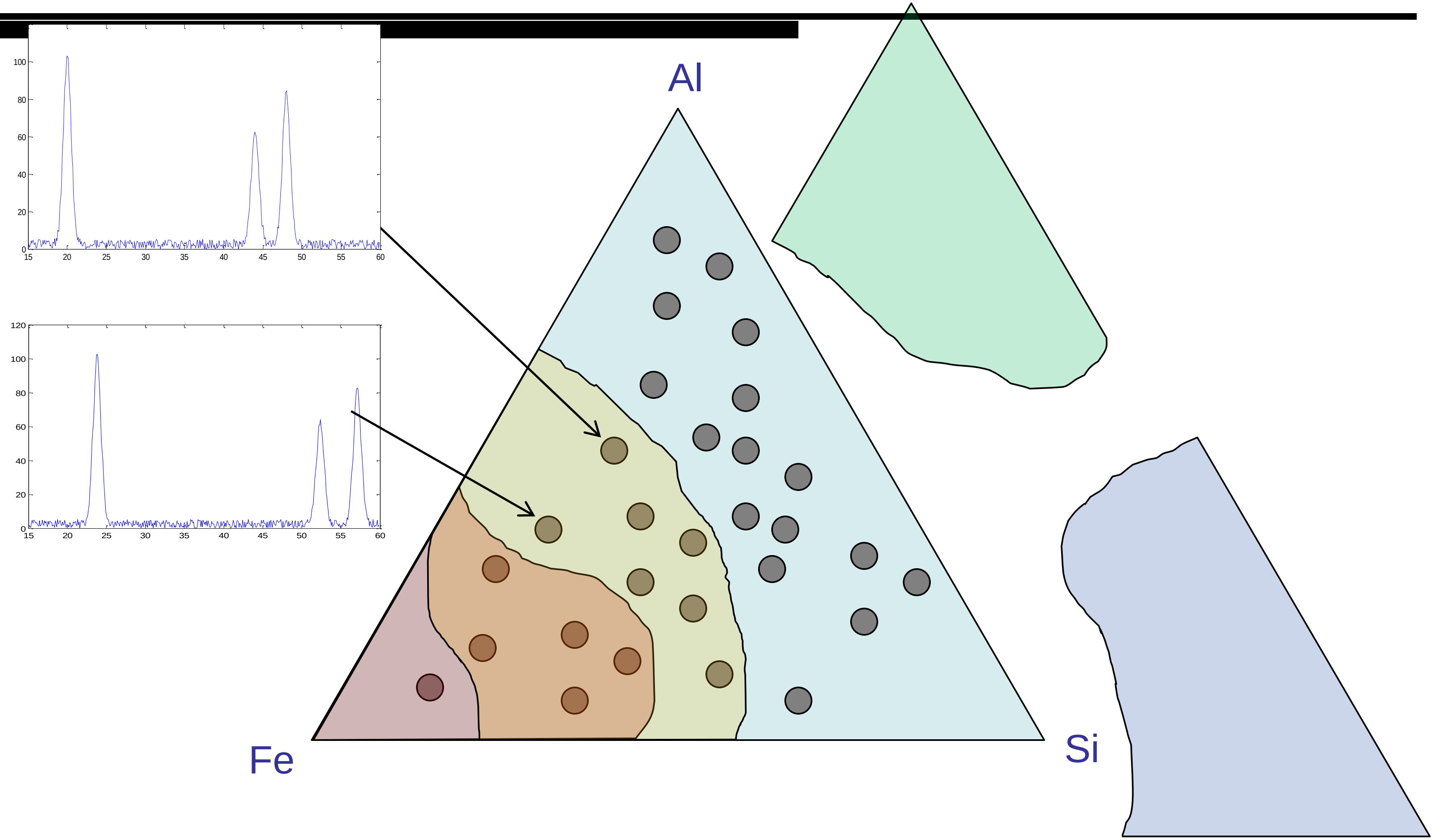
Problem Definition



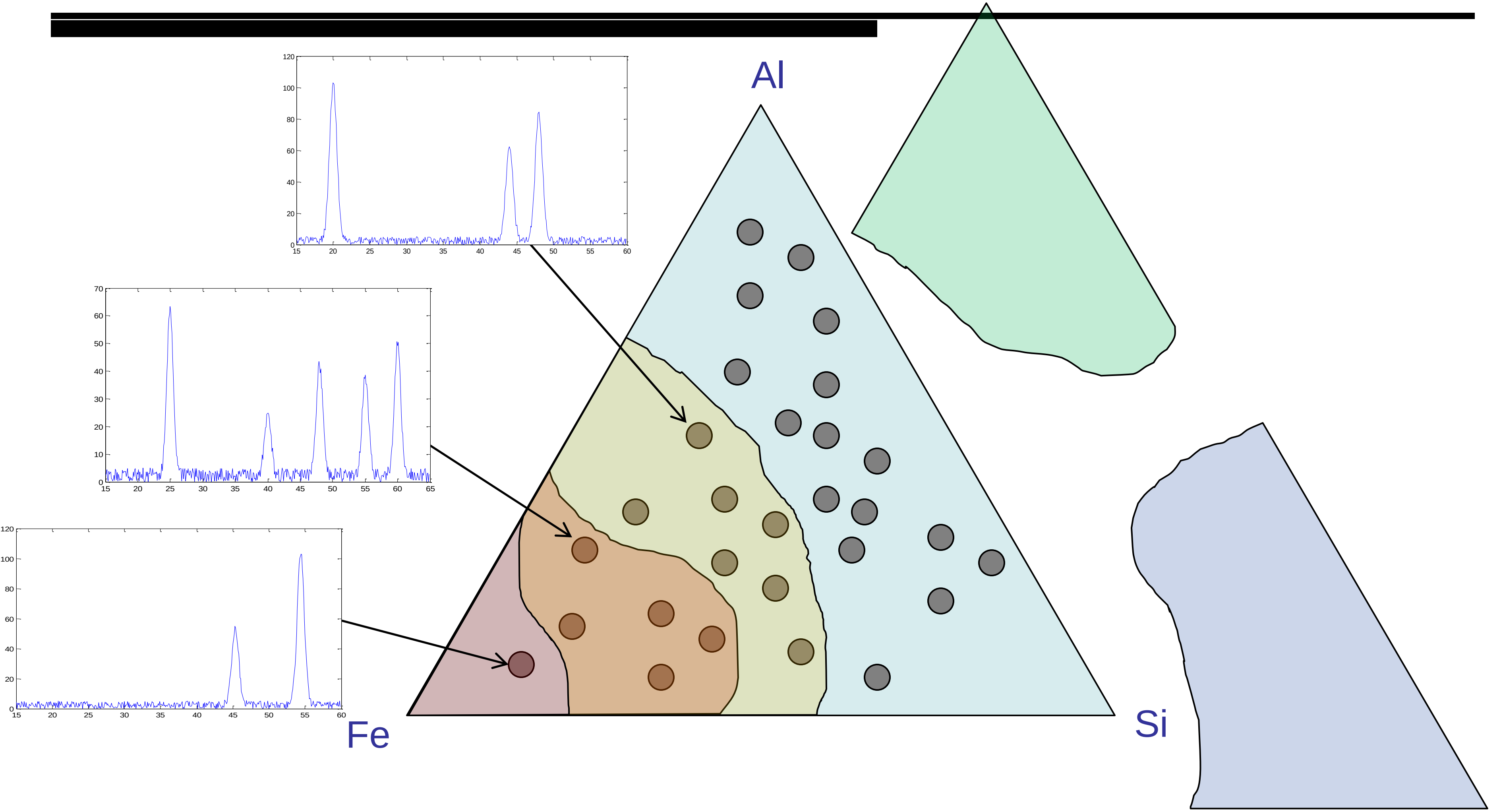
Problem Definition



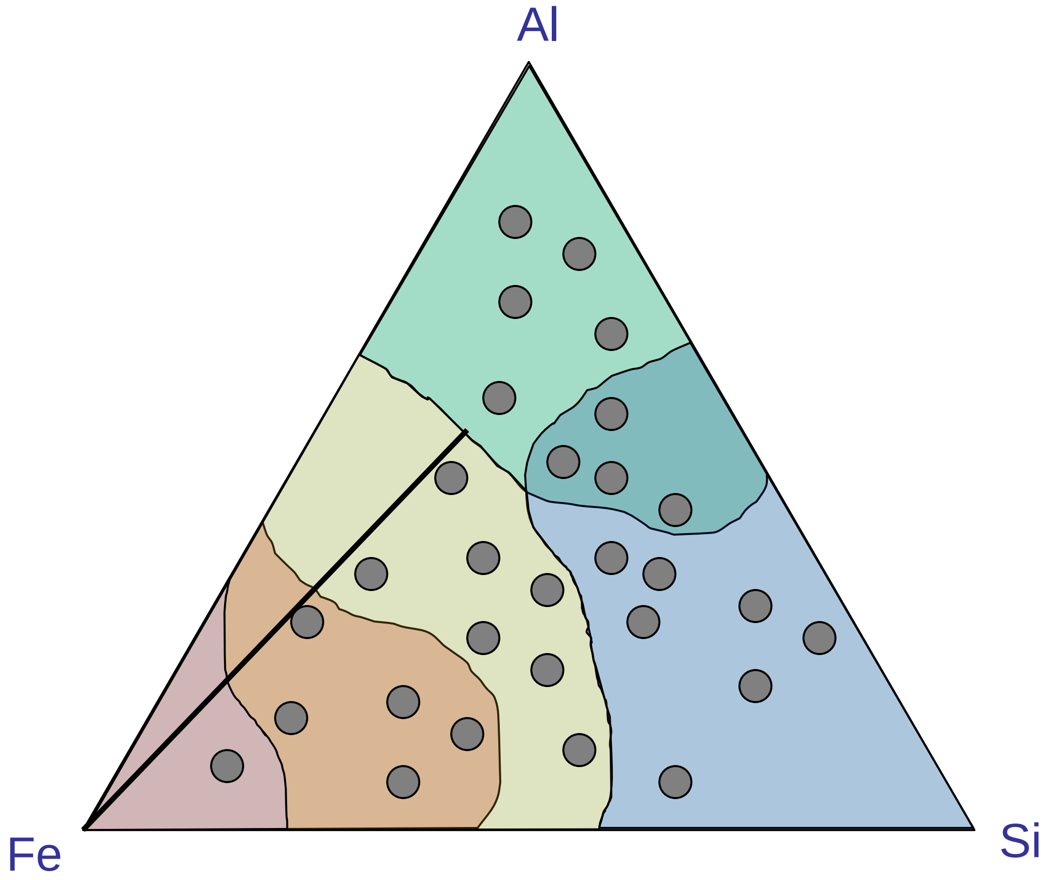
Problem Definition



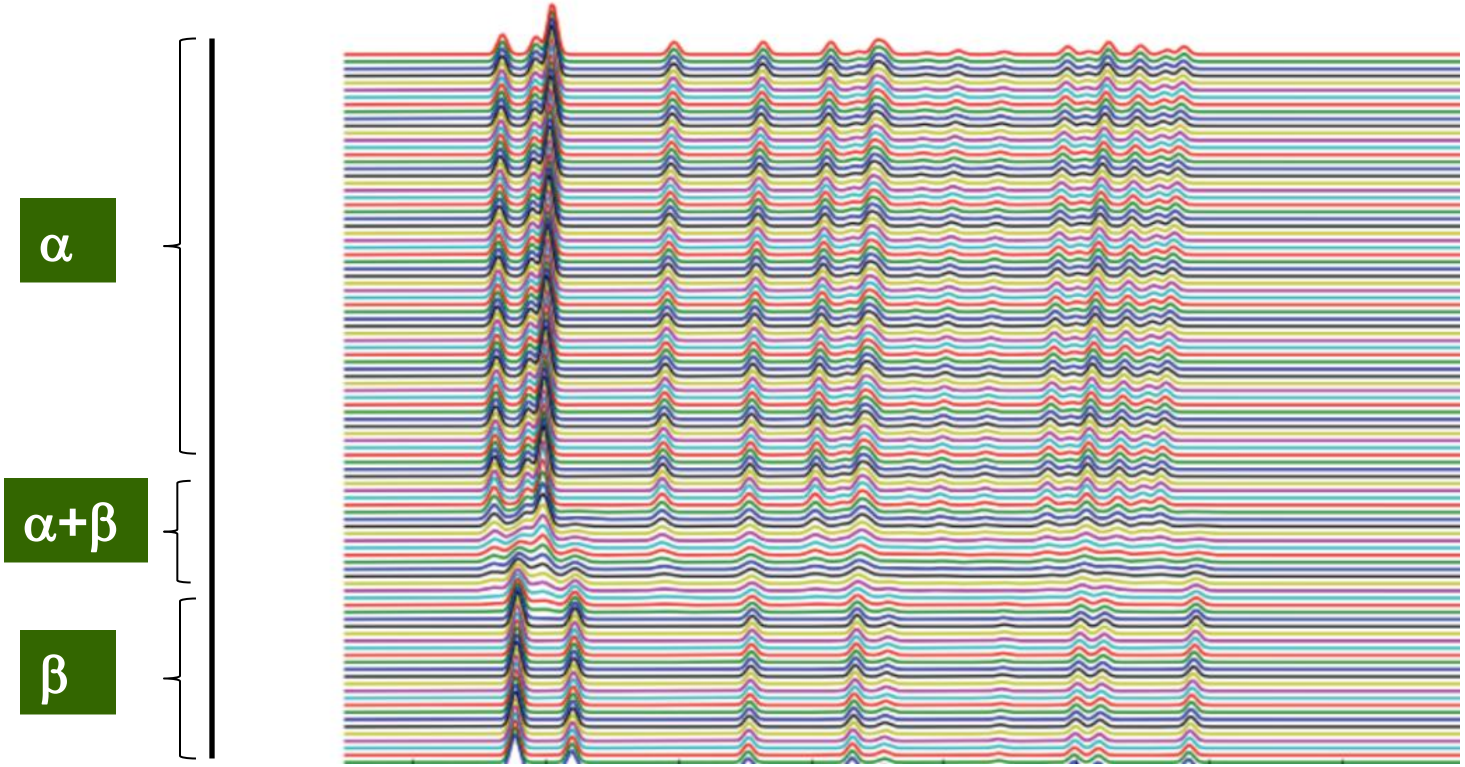
Problem Definition



Problem Definition

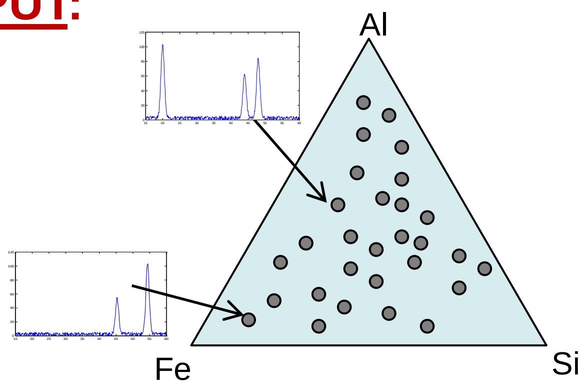


Problem Definition

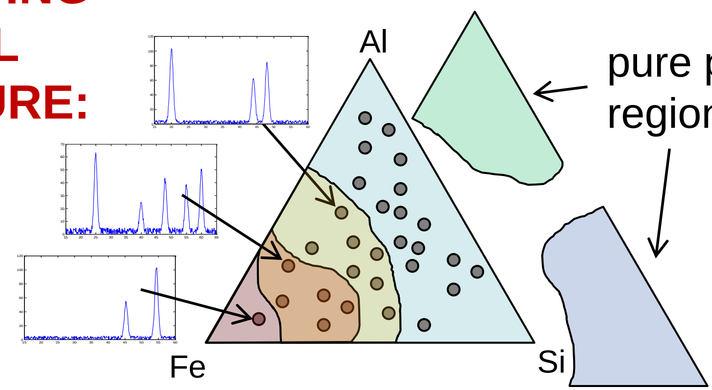


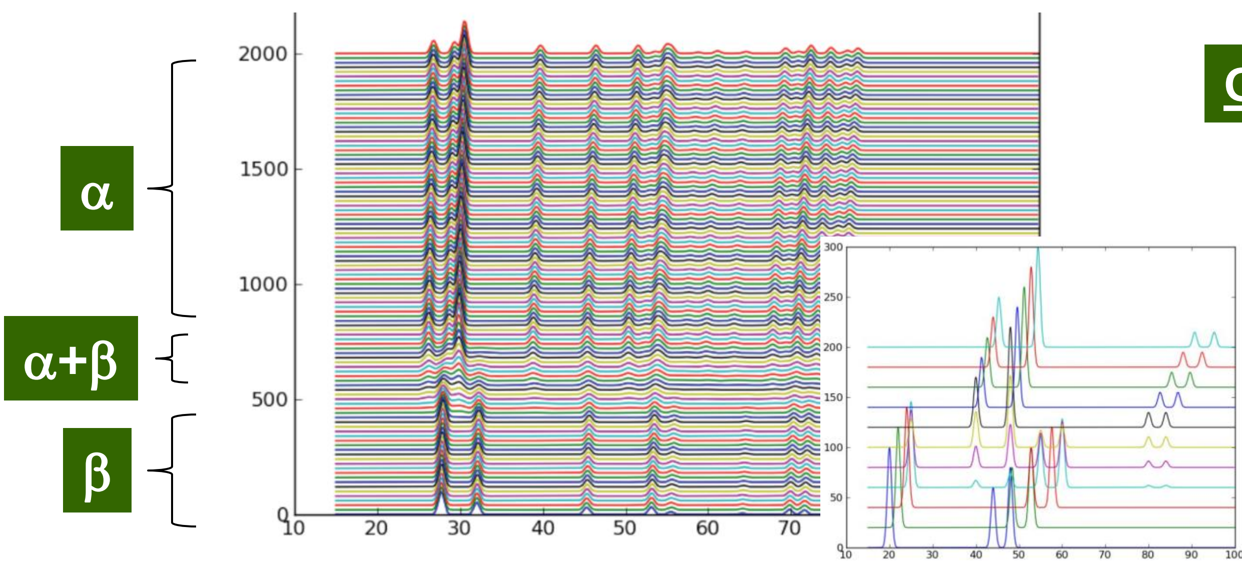
The Problem: Labeling Points with "Phase(s)"

INPUT:



UNDERLYING PHYSICAL STRUCTURE:

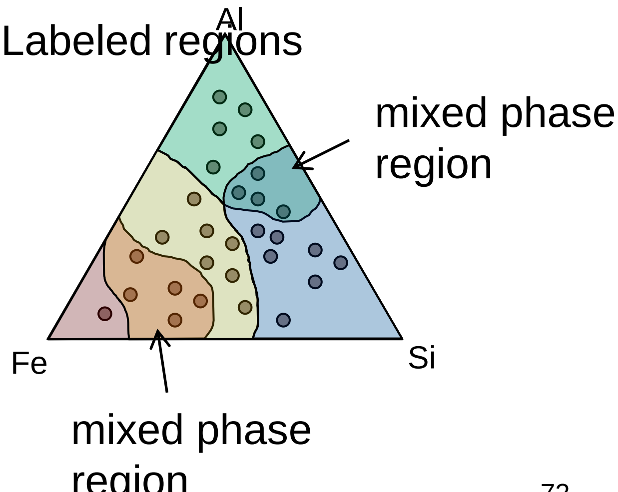




OUTPUT:

NP-hard

Labeled regions



Synthesis of Constraint Reasoning and Machine Learning Approaches for Material Discovery

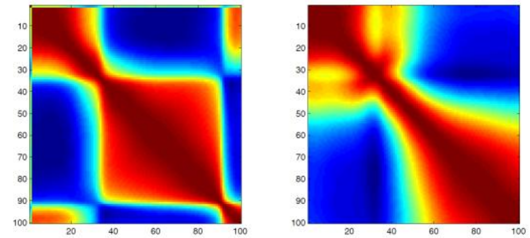
Standard Clustering Approaches:

Based on pure Machine learning techniques

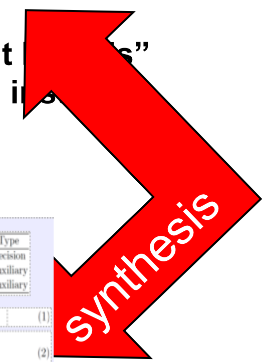
+ Good at providing a “rough”, data driven, global picture

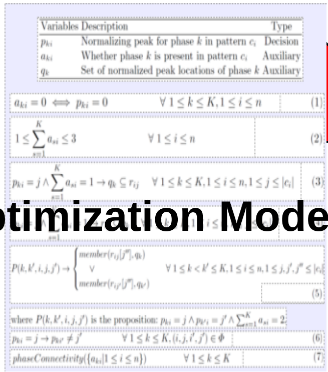
+ Can incorporate complex dependencies and global similarity structure

- Easily miss critical details (physical properties)



Two “global alignment ... s”
for a ternary problem i...





Variables	Description	Type
p_{ki}	Normalizing peak for phase k in pattern c_i	Decision
a_{ki}	Whether phase k is present in pattern c_i	Auxiliary
q_k	Set of normalized peak locations of phase k	Auxiliary

$$a_{ki} = 0 \iff p_{ki} = 0 \quad \forall\, 1 \le k \le K, 1 \le i \le n \quad (1)$$

$$1 \le \sum_{s=1}^{K} a_{si} \le 3 \quad \forall\, 1 \le i \le n \quad (2)$$

$$p_{ki} = j \wedge \sum_{s=1}^{K} a_{si} = 1 \rightarrow q_k \subseteq r_{ij} \quad \forall\, 1 \le k \le K, 1 \le i \le n, 1 \le j \le |c_i| \quad (3)$$

$$P(k,k',i,j,j') \rightarrow \begin{cases} member(r_{ij}[j''], q_k) \\ \vee \\ member(r_{ij'}[j''], q_{k'}) \end{cases} \quad \forall\, 1 \le k < k' \le K, 1 \le i \le n, 1 \le j, j', j'' \le |c_i| \quad (5)$$

where $P(k,k',i,j,j')$ is the proposition: $p_{ki} = j \wedge p_{k'i} = j' \wedge \sum_{s=1}^{K} a_{si} = 2$.

$$p_{ki} = j \rightarrow p_{ki'} \ne j' \quad \forall\, 1 \le k \le K, (i,j,i',j') \in \Phi \quad (6)$$

$$phaseConnectivity(\{a_{ki} | 1 \le i \le n\}) \quad \forall\, 1 \le k \le K \quad (7)$$

Optimization Model

Constraint reasoning and optimization

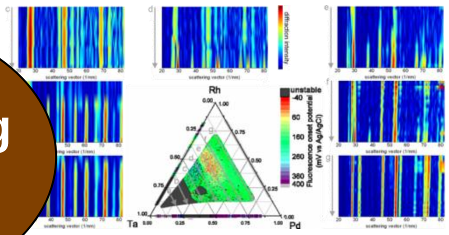
+ Great at enforcing the detailed constraints
+ Can encode and capture the “physics” behind the process

- Do not scale up

What's New: Solving it "Properly" Requires…

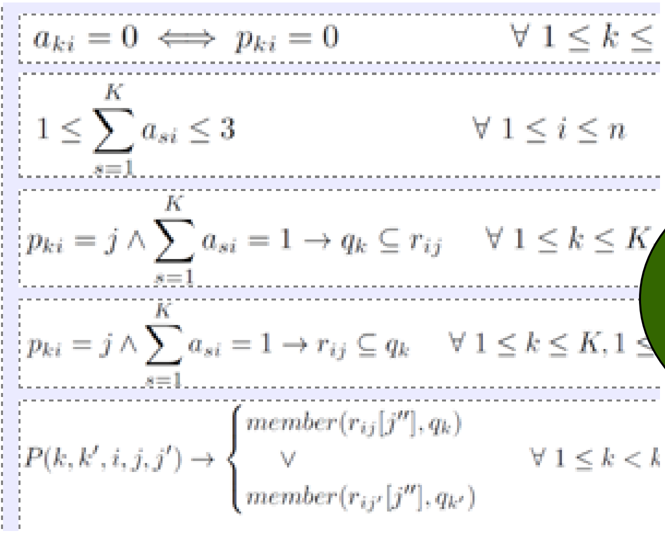
… a robust, *physically meaningful*, scalable, automated solution method that combines:

Underlying <u>Physics</u>

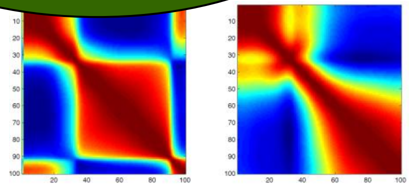


A language for <u>Constraints</u> enforcing "local details"

<u>Machine Learning</u> for a "global data-driven view"



Constraint Programming model



Similarity "Kernels" & Clustering

Significantly outperforming previous methods.

Outline

- Computational Sustainability Research @ Cornell
  - ICS Research team
  - A sample of Interdisciplinary Research Projects (IRPs)
  - Main computational themes
- The Institute for Computational Sustainability
- Building a community in Computational Sustainability
- Conclusions

Deep Research Challenges posed by Sustainability → Transformative Computer Science Research

Design of policies to manage natural resources translating into large-scale optimization and learning problems, combining a mixture of discrete and continuous effects, in a highly dynamic and uncertain environment → increasing levels of complexity

Transformative Computer Science Research:
Driven by Deep Research Challenges posed by Sustainability

Design of policies to manage natural resources translating into large-scale optimization and learning problems, combining a mixture of discrete and continuous effects, in a highly dynamic and uncertain environment → increasing levels of complexity

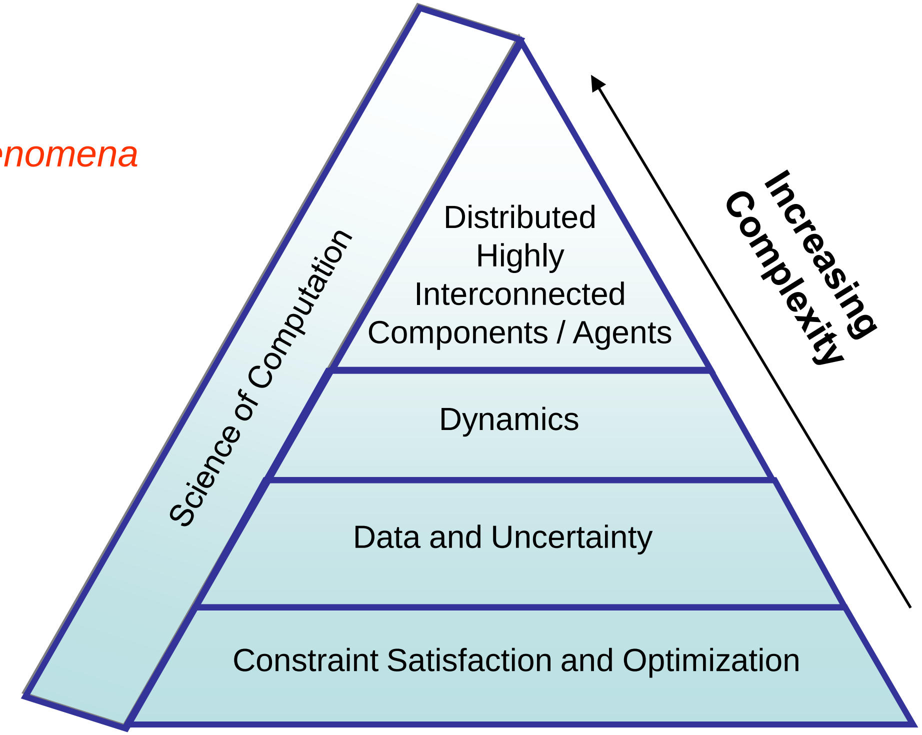
Study computational problems as natural phenomena
→ **Science of Computation**

Many highly interconnected components;
→ **From Centralized to Distributed: multi agent systems**

Complex dynamics and Multiple scales
→ **From Statics to Dynamics: Complex systems and dynamical systems**

Large-scale data and uncertainty
→ **Machine Learning, Statistical Modeling**

Complex decision and optimization problems
→ **Constraint Reasoning and Optimization**



Complexity levels in Computational Sustainability Problems

Transformative Computer Science Research:
Driven by Deep Research Challenges posed by Sustainability

Design of policies to manage natural resources translating into large-scale optimization and learning problems, combining a mixture of discrete and continuous effects, in a highly dynamic and uncertain environment

→ increasing levels of complexity

Study computational problems as natural phenomena

→ **Science of Computation**

Many highly interconnected components;

→ **From Centralized to Distributed: multi agent systems**

Complex dynamics and Multiple time scales

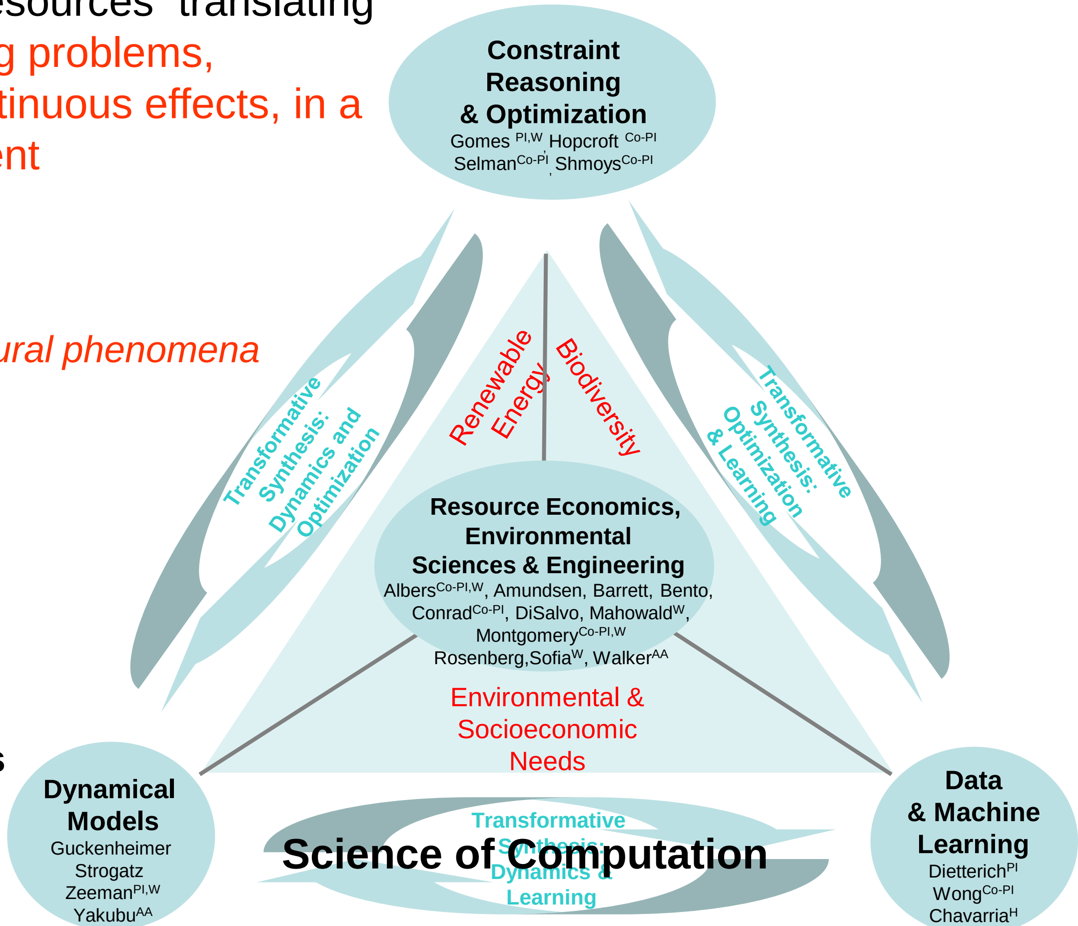
→ **From Statics to Dynamics: Complex systems and dynamical systems**

Large-scale data and uncertainty

→ **Machine Learning, Statistical Modeling**

Complex decision and optimization problems

→ **Constraint Reasoning and Optimization**



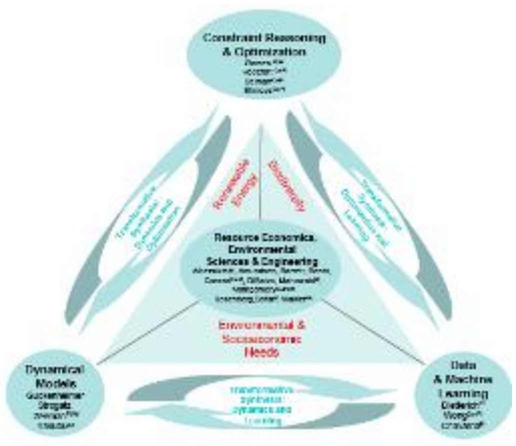
Science of Computation

Outline

- Computational Sustainability Research at the Institute for Computational Sustainability @ Cornell
  - Main computational themes @ ICS
  - A sample of Interdisciplinary Research Projects (IRPs)
- The Institute for Computational Sustainability
- Building a community
- Conclusions

Institute Activities

Research



Coordinating transformative synthesis collaborations

Interdisciplinary Research Projects (IRPs)

Building Research Community

Conference & Workshops

Web Portal

Seminars

Host visiting Scientists

External Collaborations

ICS

Education

Postdocs

Doctoral students

Honors projects

Research seminar series

Summer REU program targeting minority students

Computational Sustainability courses

Outreach

K-12 Activities

Citizen Science Lab. Of Ornithology

Conservation Fund

Cornell Center for a Sustainable Future

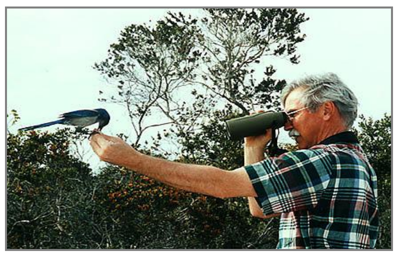
Cornell Cooperative Extension

Science Exhibits

The Nature Conservancy

Outreach An Example: Citizen Science at the Cornell Lab. Of Ornithology



- ***Increase scientific knowledge***

  Gather meaningful data to answer large-scale research questions
- ***Increase scientific literacy***

  Enable participants to experience the process of scientific investigation and develop problem-solving skills
- ***Increase conservation action***

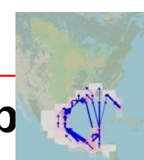
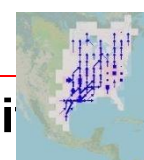
  Apply results to science-based conservation efforts

Examples of research and outreach outputs:,

CLO: Track record of influencing conservation science.

Online Research Kit

...project at the Lab ...ni...gy at Cornell ...in... - from research labs to backyard ...ne... unts birds to contribute to research.

Maps with Species Distributions and Migration Flows

Conservation Plan Partners in Flight International (Rosenberg, Co-Chair)

Outline

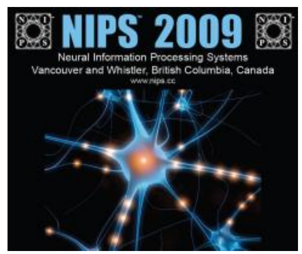
- Computational Sustainability Research at the Institute for Computational Sustainability @ Cornell
  - Main computational themes @ ICS
  - A sample of Interdisciplinary Research Projects (IRPs)
- The Institute for Computational Sustainability
- Building a community
- Conclusions

Computational Sustainability Community Conferences, Workshops, Discussion Groups

Blogs:

CompSust'09: Cornell, 2009

Over 225 international researchers from several disciplines and institutions (universities, labs, government)



- NIPS-09 Mini Symposium: Machine Learning for Computational Sustainability

International Workshop on Constraint Reasoning and Optimization for Computational Sustainability

- **CROCS**-09 at CP-09
- CROCS at CPAIOR-10
- CROCS at CP-10

CompSust'10: MIT, 2010

Chairs: Brian Williams, Tom Dietterich, Carla Gomes

Over 150 participants and lots of topics

Summary

Computational Sustainability

has great potential to advance the state of the art of computer science and related disciplines and with unique societal benefits!

The End ☺!

Thank you!