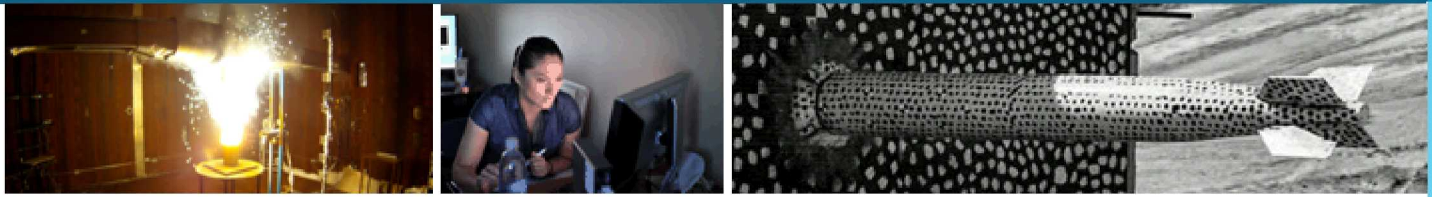
SAND2019-0448C

Working with Harvey Greenberg: Computational Biology, Water Security and More!



William E. Hart

Sandia National Laboratories
wehart@sandia.gov

Working With Harvey Greenberg

2000 – Harvey First Visits Sandia

2004
- Opportunities for combinatorial optimization in computational biology
- HJ Greenberg, WE Hart, G Lancia

2004
- A multiple-objective analysis of sensor placement optimization in water networks
- JP Watson, HJ Greenberg, WE Hart

2006
- Robust optimization of contaminant sensor placement for community water systems
- RD Carr, HJ Greenberg, WE Hart, G Konjevod, E Lauer, H Lin, T Morrison, CA Phillips

2007
- Enumeration of Near-Optimal Solutions for Peptide Docking Problems
- WE Hart, RD Carr, HJ Greenberg, ME Hart, R Forrester, D Roe

2007
- LDRD final report: robust analysis of large-scale combinatorial applications.
- RD Carr, T Morrison, WE Hart, NL Benavides, HJ Greenberg, JP Watson, CA Phillips

Working With Harvey Greenberg

2000 – Harvey First Visits Sandia

2004
- Opportunities for combinatorial optimization in computational biology
- HJ Greenberg, WE Hart, G Lancia

2004
- A multiple-objective analysis of sensor placement optimization in water networks
- JP Watson, HJ Greenberg, WE Hart

2006
- Robust optimization of contaminant sensor placement for community water systems
- RD Carr, HJ Greenberg, WE Hart, G Konjevod, E Lauer, H Lin, T Morrison, CA Phillips

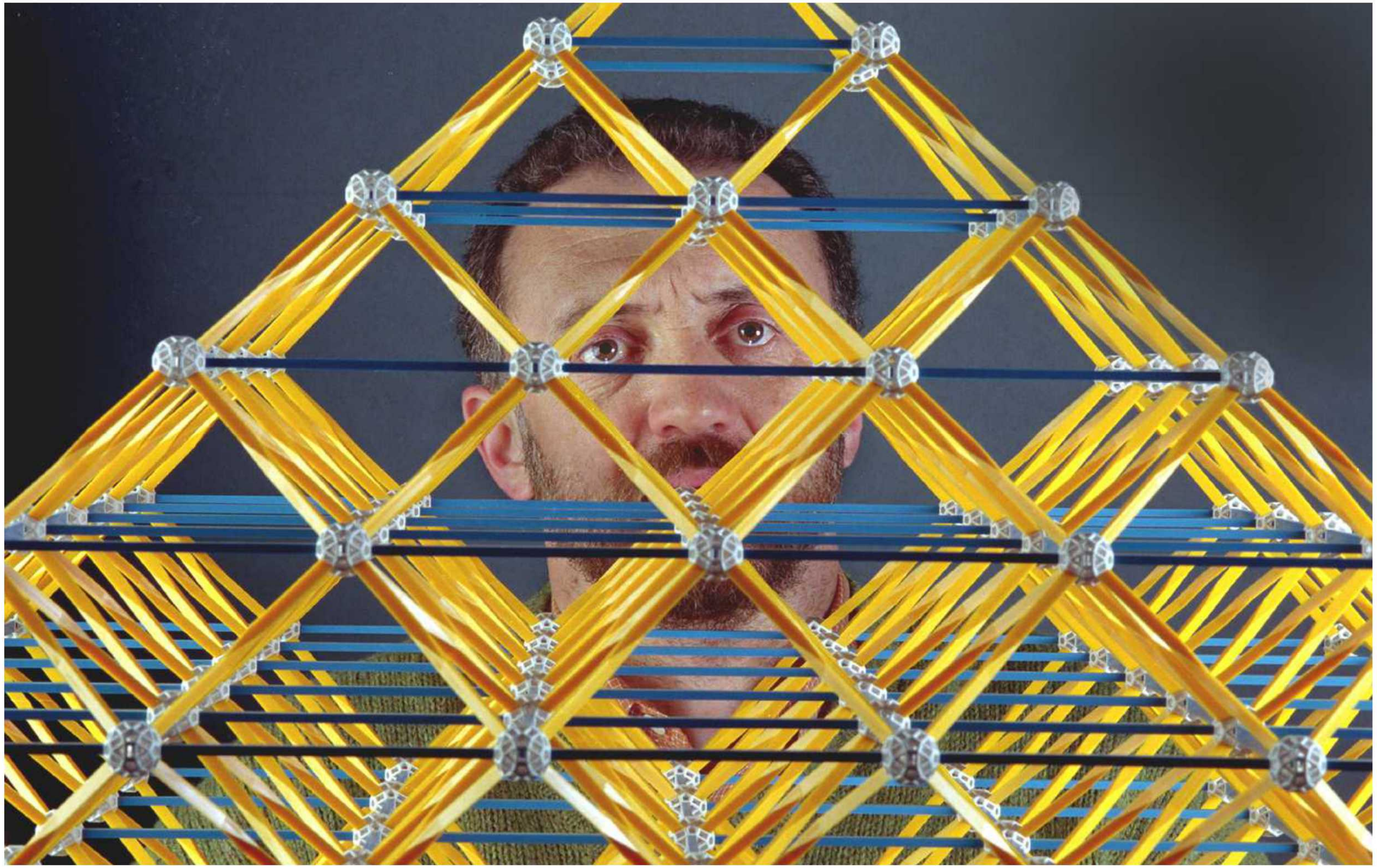
2007
- Enumeration of Near-Optimal Solutions for Peptide Docking Problems
- WE Hart, RD Carr, HJ Greenberg, ME Hart, R Forrester, D Roe

2007
- LDRD final report: robust analysis of large-scale combinatorial applications.
- RD Carr, T Morrison, WE Hart, NL Benavides, HJ Greenberg, JP Watson, CA Phillips

OR Methods for Computational Biology

Learning from Sorin Istrail

Immersed in Computational Biology in Sorin's old office …

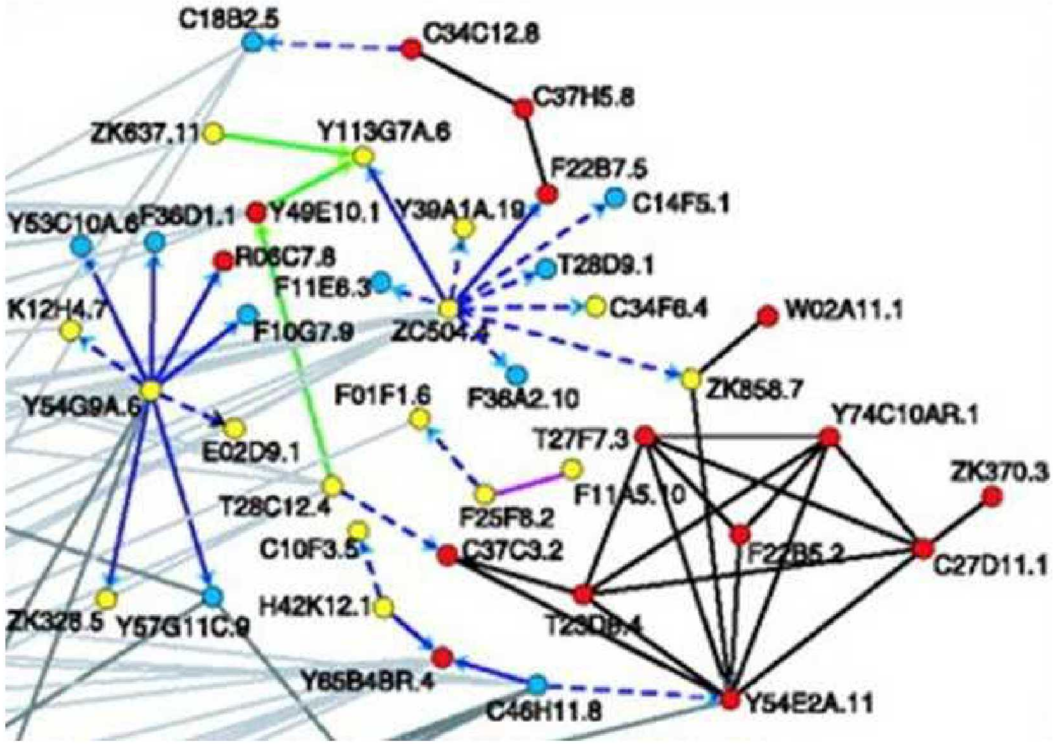


Example: Analyzing Protein-Protein Interactions

Protein-protein interactions are essential to virtually all cellular processes

- Proteins are macro-molecules composed of a linear chain of amino acid residues
- Proteins folds into well-defined 3D structures that determine their role in cellular processes

Protein-protein interactions provide insight into protein function



Protein-Peptide Docking

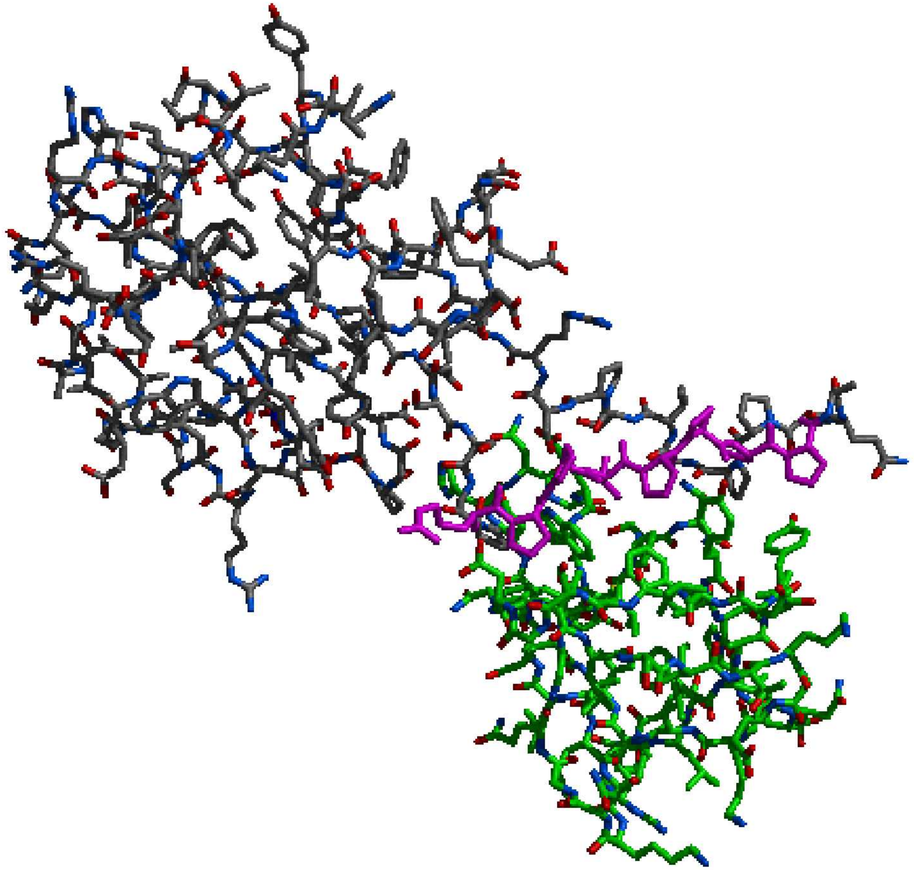
Idea: study protein-protein interactions through protein-peptide docking

Many protein-protein interactions are mediated by modular domains

- PDZ, SH3, SH2, WW, PTB, FHA

These domains often bind to a linear stretch of the binding partner

Impact: Peptide libraries can identify binding sequences that can be used to search the genome for binding partners



Complex of the SH3 domain from tyrosine protein kinase (green) complexed with P38(grey), with linear portion highlighted (magenta)

Computational Phage-Display

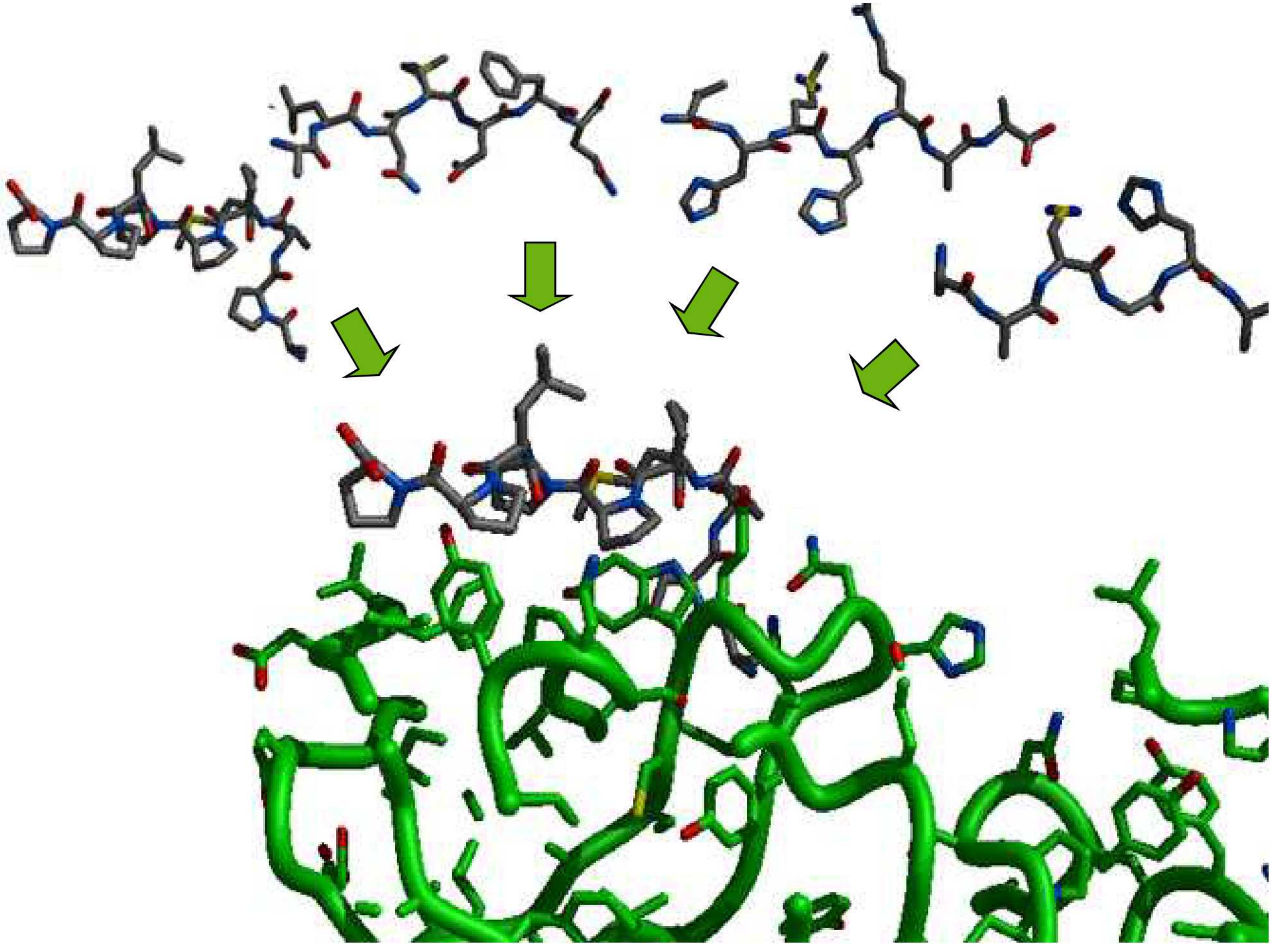
Idea: predict the peptide sequence that is most likely to bind

- Predict consensus sequences

Impact:

- Provide the same type of information as experimental methods

- Can be used to design a reduced “focused” phage-display library (minimizes experimental costs)



Computational OR Approach

Three problems with a single mathematical representation:

1. Predict the structure of a peptide
2. Predict the best structure of a peptide allowing for a set of possible amino acids for each sidechain
3. Predict the best amino acids for the sidechains in a peptide

Assumptions

1. Side-chain structures can be well-captured by rotamer libraries
   - These provide a discretization of the design space
2. The peptide backbone is well-constrained
   - The backbone is often assumed fixed

Harvey’s Impact

OR in Computational Biology
- INFORMS J. Computing Survey
- Center for Computational and Mathematical Biology

Computational Phage Display
- More in Dick’s talk

Enumeration strategies that identify near-optimal solutions
- Can be use to limit the scope of phage-display experiments
  - Only include amino acids at sites where they appear in near optimal solutions

- Can be used to identify peptide docking candidates
  - Used to define a consensus matrix
  - Can quickly scan the genome for binding partners

Working With Harvey Greenberg

2000 – Harvey First Visits Sandia

2004
- Opportunities for combinatorial optimization in computational biology
- HJ Greenberg, WE Hart, G Lancia

2004
- A multiple-objective analysis of sensor placement optimization in water networks
- JP Watson, HJ Greenberg, WE Hart

2006
- Robust optimization of contaminant sensor placement for community water systems
- RD Carr, HJ Greenberg, WE Hart, G Konjevod, E Lauer, H Lin, T Morrison, CA Phillips

2007
- Enumeration of Near-Optimal Solutions for Peptide Docking Problems
- WE Hart, RD Carr, HJ Greenberg, ME Hart, R Forrester, D Roe

2007
- LDRD final report: robust analysis of large-scale combinatorial applications.
- RD Carr, T Morrison, WE Hart, NL Benavides, HJ Greenberg, JP Watson, CA Phillips

OR Methods for Water Security

Water Security

Universal Vulnerabilities in Water Systems

- Plant access
- Source Water
- Water storage
- Water distribution

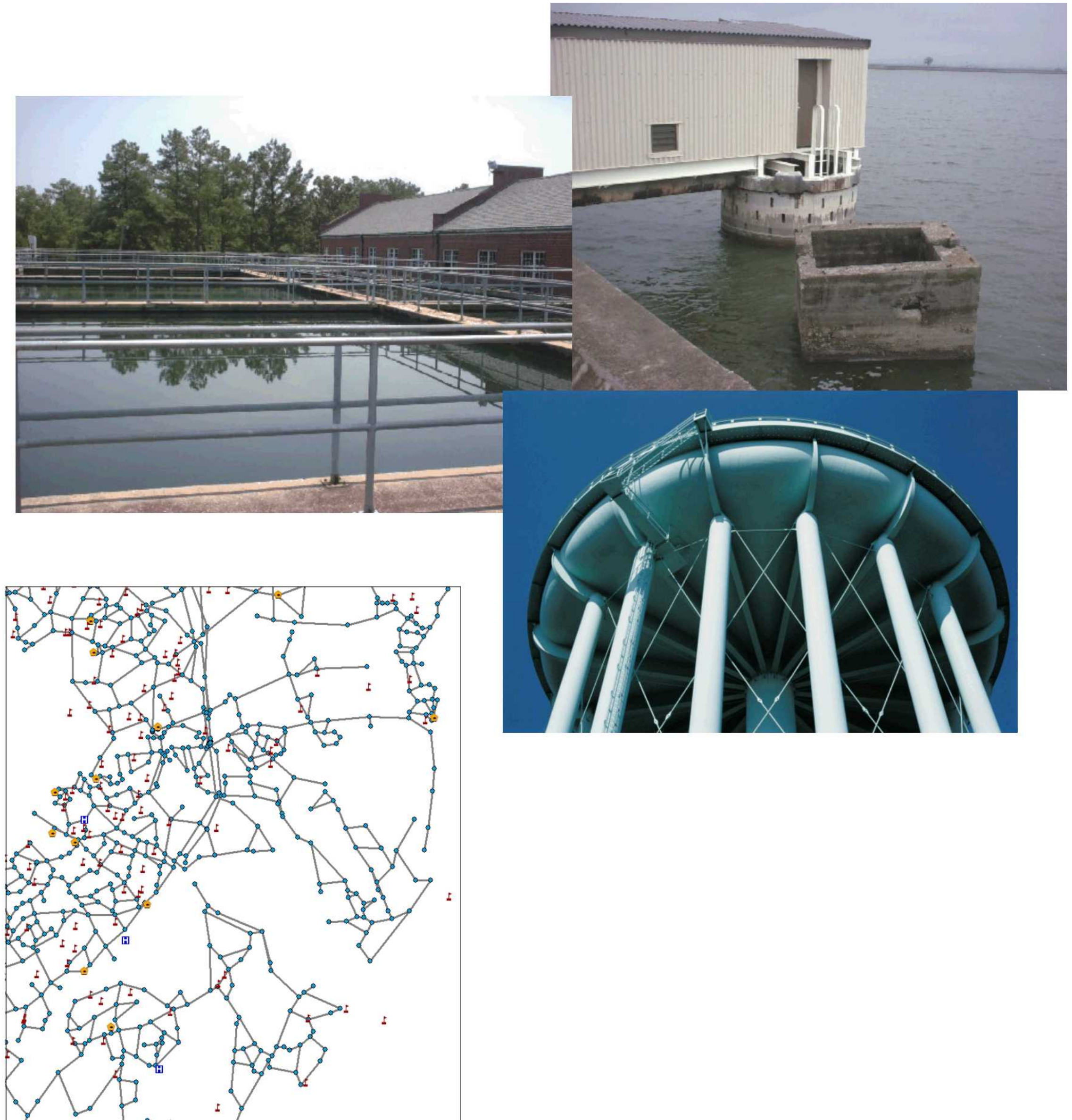
Challenge: Minimizing risk of contamination threats in municipal water networks

Goal: Develop early warning system

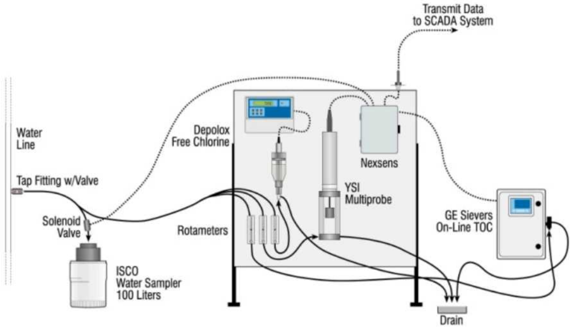
- Sensor network to quickly detect contaminants
- Focus on protection of human populations

Idea: Place sensors at junctions

- E.g. utility pipes, schools, hospitals

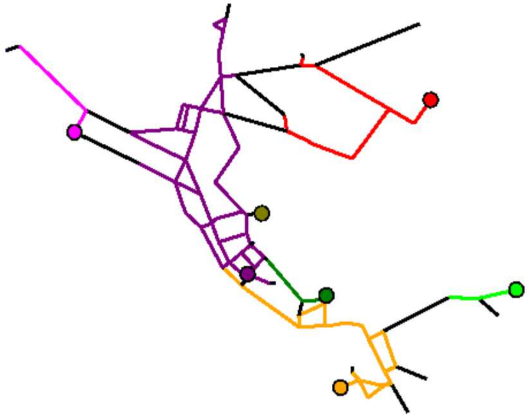


Sensor Placement Optimization



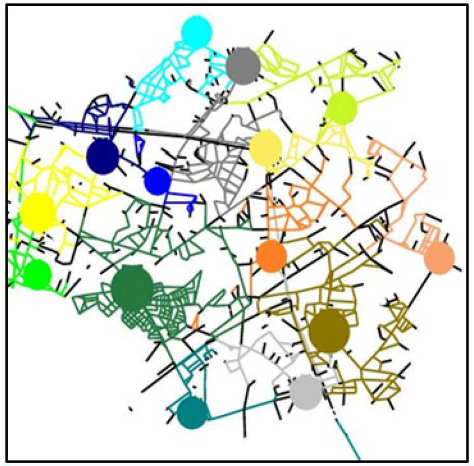
Sensor Characteristics

Contamination Incidents



Water Networks

Optimization
Solvers



Sensor Placements

Impact Summary and
Confidence Assessment

Integer Programming for Sensor Placement

IPs can be used to model sensor placement for water security

- Berry et al (2003, 2006); Watson et al (2004)

Objective: $\boxed{\sum_{a\in A}\sum_{i\in L}\alpha_a w_{ai} x_{ai}}$

- α – contamination likelihood
- w – contamination impact
- x – witness variable
- s – sensor placement variable

IP model can capture different objectives/networks

Can be solved with COTS software on 64-bit computers

$$
\begin{array}{ll}
\text{minimize } \sum_{a\in A}\sum_{i\in L}\alpha_a w_{ai} x_{ai} & \\
s.t. & \\
\sum_{i\in L} x_{ai} = 1 & \forall a \in A \\
x_{ai} \le s_i & \forall a \in A, i \in L \\
\sum_{i\in L} s_i \le S_{\max} & \\
s_i \in \{0,1\} & \\
0 \le x_{ai} \le 1 & \forall a \in A, i \in L
\end{array}
$$

NOTE: This is a p-median problem!

Harvey’s Impact

Rethinking core modeling assumptions…

1. Minimize risk measures instead of expected impacts
   - Value at Risk (VaR) and Conditional Value at Risk (CVaR) formulations
   - Evaluated risk-performance trade-offs

2. Our data is imprecise
   - E.g. We make simplifying assumptions about who is drinking water
   - We need to account for data errors in our sensor network designs

3. Assessing trade-offs between competing objectives
   - There are many potential objectives
     - E.g. min cost, min # deaths, min extent of contamination, min time of detection
   - Demonstrated goal programming approaches to find trade-offs
   - Led to related work that was awarded a best paper award for the J Water Works Assoc.

Working With Harvey Greenberg

2000 – Harvey First Visits Sandia

2004
- Opportunities for combinatorial optimization in computational biology
- HJ Greenberg, WE Hart, G Lancia

2004
- A multiple-objective analysis of sensor placement optimization in water networks
- JP Watson, HJ Greenberg, WE Hart

2006
- Robust optimization of contaminant sensor placement for community water systems
- RD Carr, HJ Greenberg, WE Hart, G Konjevod, E Lauer, H Lin, T Morrison, CA Phillips

2007
- Enumeration of Near-Optimal Solutions for Peptide Docking Problems
- WE Hart, RD Carr, HJ Greenberg, ME Hart, R Forrester, D Roe

2007
- LDRD final report: robust analysis of large-scale combinatorial applications.
- RD Carr, T Morrison, WE Hart, NL Benavides, HJ Greenberg, JP Watson, CA Phillips

OR Modeling Environments

Compiling an Optimization Problem

An optimization problem needs to be "compiled" into a form that a solver can understand

Examples:

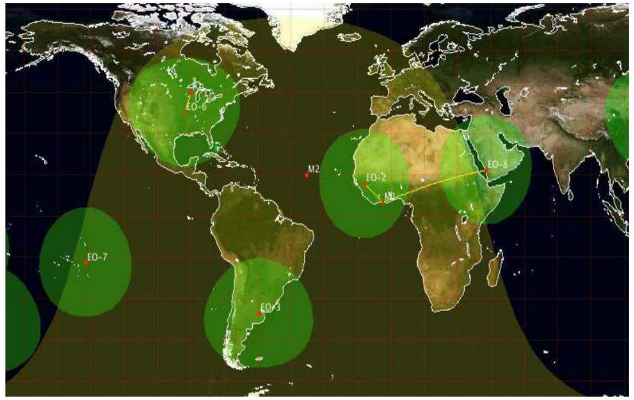
1. A code or script that computes objectives, constraints, derivatives, …
    - This includes a subroutine that does this computation
2. Matrices and vectors used in a standard form (e.g. for linear programming)

$$\boxed{\begin{array}{ll} \min_x & c^T x \\ & Ax = b \\ & x \geq 0 \end{array}}$$

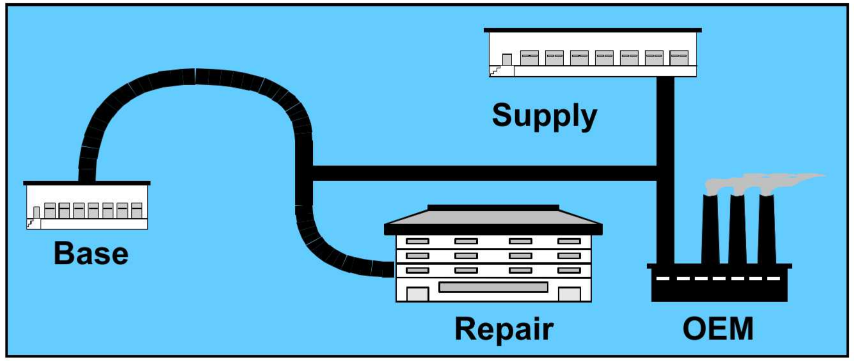
3. A general representation of the problem expressed in an Algebraic Modeling Language (AML)

Algebraic Modeling Languages

AMLs are key to solving complex national security applications.

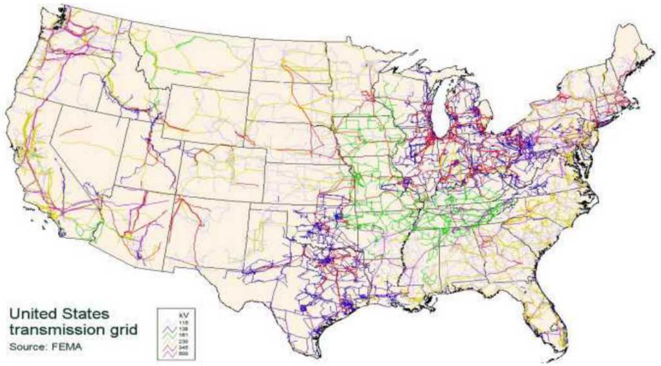


Satellite Scheduling



Military Logistics

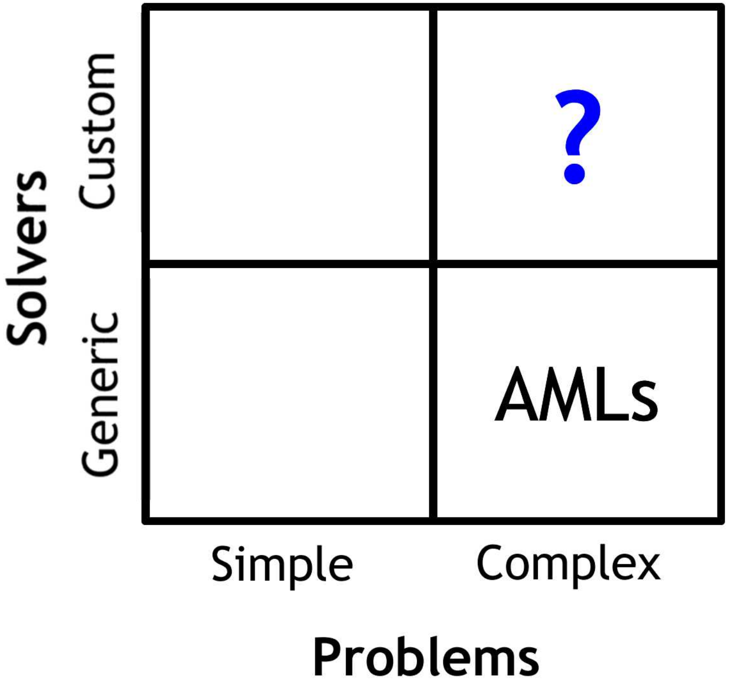
Water Security



United States transmission grid
Source: FEMA

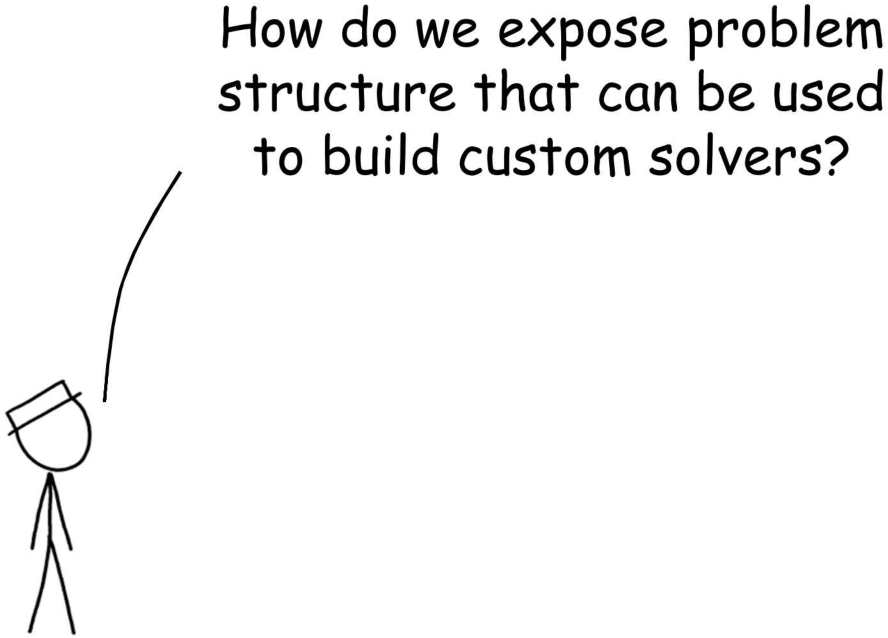
Power Grid Planning

We Need AMLs for Both Research and Applications



Solvers	Simple	Complex
Custom		?
Generic		AMLs

Problems

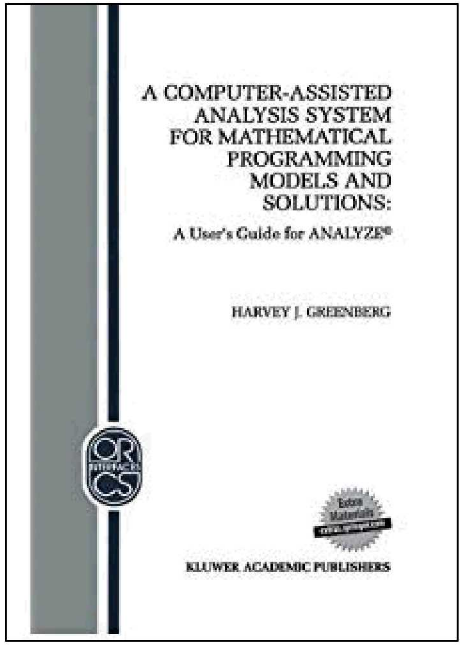


Harvey’s Impact

My conversations with Harvey about this challenge inspired the development of Pyomo.

Harvey developed Computer-Assisted Analysis in the 1970s and 80s, creating an artificially intelligent environment for analyzing mathematical programming models and their results.

Harvey was awarded the first ICS Prize for "research excellence in the interfaces between operations research and computer science" in 1986, notably for his software system, ANALYZE.



A COMPUTER-ASSISTED
ANALYSIS SYSTEM
FOR MATHEMATICAL
PROGRAMMING
MODELS AND
SOLUTIONS:
A User's Guide for ANALYZE®

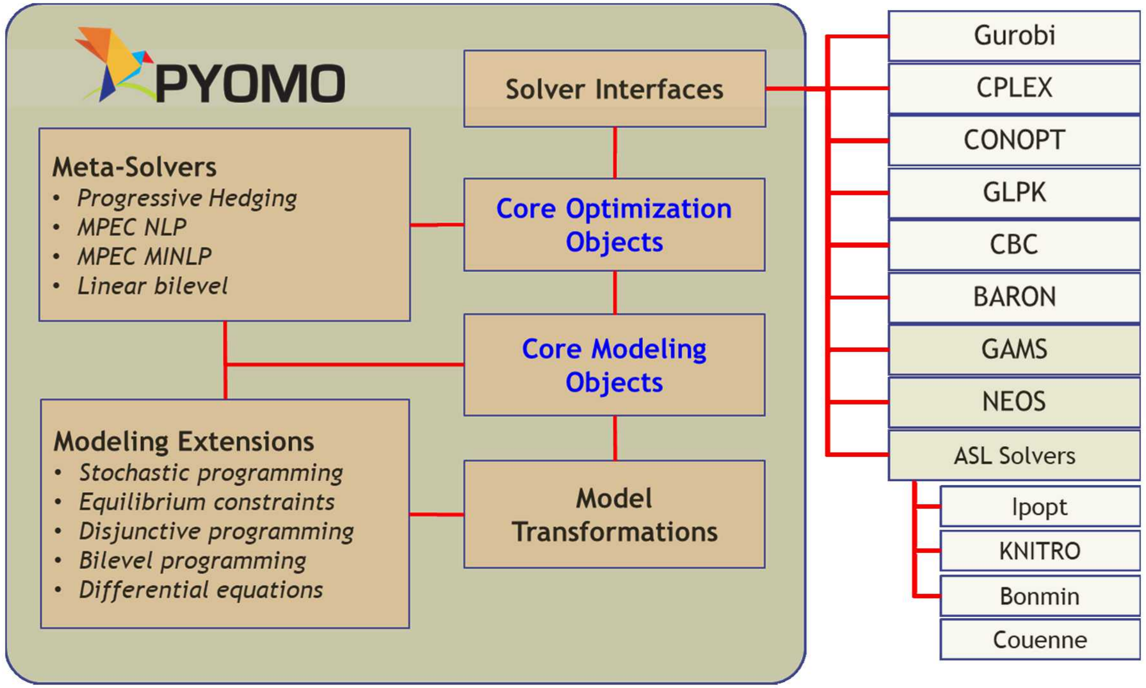
HARVEY J. GREENBERG

KLUWER ACADEMIC PUBLISHERS

Pyomo

Pyomo is an open-source AML

- Built on Python, a full-featured programming language with a very clean syntax
- Can model a very diverse set of problems
  - Stochastic programs, bilevel programs, MPECs, GDPs, DAE, etc
- Can express hierarchical model structure
- User-control of model transformations
- Extensible object-oriented software design



PYOMO

Solver Interfaces

Meta-Solvers
- *Progressive Hedging*
- *MPEC NLP*
- *MPEC MINLP*
- *Linear bilevel*

Core Optimization Objects

Core Modeling Objects

Modeling Extensions
- *Stochastic programming*
- *Equilibrium constraints*
- *Disjunctive programming*
- *Bilevel programming*
- *Differential equations*

Model Transformations

Gurobi
CPLEX
CONOPT
GLPK
CBC
BARON
GAMS
NEOS
ASL Solvers
  Ipopt
  KNITRO
  Bonmin
  Couenne

Pyomo Highlights

- Pyomo was selected as one of the 100 winners of the 2016 R&D100 Awards!
- >75,000 downloads in 2018
- >24,000 downloads in 2017 of Pyomo book chapters

On a Personal Note…