Sustainability is an uncharted territory for computer scientists, highly interdisciplinary area.

To ensure that the field of computational sustainability becomes self-sustaining

Critical mass and visibility that cannot be achieved with piece-meal research efforts.

2nd Expedition (2016)
CompSustNet: Large-Scale Research Network
Expanding the Horizons of Computational Sustainability
Computational Sustainability

New interdisciplinary field that aims to advance computational methods for 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:

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

- Stated the urgency of policies for sustainable development.

**Gro Brundtland**
Norwegian Prime Minister
Chair of WCED

Sustainability: Interlinked environment, economic, and social

Sustainable Development encompasses balancing

• environmental,
• economic, and
• societal needs.

Ultimate goal of Sustainable Development

HUMAN WELL-BEING of current and future generations.
Sustainability problems
*unique in scale and complexity*

Smart Power Grid:
Complex Digital Ecosystem

Key challenge:
How to effectively establish the necessary large-scale interdisciplinary projects and collaborations

CompSustNet
Large Research Network for Expanding the Horizons of Computational Sustainability
CompSustNet: Large-scale Research Computational Sustainability Network

12 US Academic Institutions
Collaborators:
10 US Organizations (Gov and NGOs) and 16 International Organizations (Academic Institutions, Gov, and NGOs)
12 US Academic Institutions
Collaborators:
10 US Organizations (Gov and NGOs) and 16 International Organizations (Academic Institutions, Gov, and NGOs)
12 US Academic Institutions
Collaborators:
10 US Organizations (Gov and NGOs)
and 16 International Organizations
(Academic Institutions, Gov and NGOs)
Total: 38 Institutions and ~80 People
Strategically Partnering with Key Centers of Excellence
(providing sustainability knowledge and infrastructure)

- The Nature Conservancy
- Secretary of Environment, Quito, Ecuador
- NASA Remote Sensing
- US Geological Survey
- ILRI International Livestock Research Institute (Kenya)
- Trans African Hydro Meteorological Observatory
- Join Center for Artificial Photosynthesis
- The National Institute of Standards and Technology
- Materials Science
- Cornell High Energy Synchrotron Source
- Cornell University Laboratory for Energy Systems Analysis
- Princeton University
- The Atkinson Center for a Sustainable Future
- University of Southern California
- Stanford University
- University of California, Davis
- University of Massachusetts Amherst
- University of Virginia
- University of Utah
- University of Illinois
- University of Washington
- University of North Carolina
- University of Wisconsin-Madison
- University of Michigan
- University of Chicago
- University of Minnesota
- University of Texas at Austin
- University of California, Berkeley
- University of California, Los Angeles
- University of California, San Diego
Core sustainability themes:

1. Biodiversity and Conservation,
2. Balancing Environmental and Socio-economic Needs, and
3. Energy and Renewable Resources.

Main computational thrusts:

1. Big data and Machine Learning,
2. Simulation, Optimization, and Dynamical Models,

Wide range of sustainability applications:

To build a **coherent and cohesive computer science sub-field** it’s paramount to **go deep** in terms of **Cross-Cutting Core Computational Problems** and apply them to different sustainability areas.
Sample of Computational Sustainability Interdisciplinary Research Projects (IRPs)

Cross-Cutting Computational Themes

Community Building

Conclusions
Fundamental question in biodiversity research: How different species are distributed across landscapes over time.
Sensing biodiversity: A variety of sensors

Several computational data challenges

- Global-regional coverage
- Regional-local coverage

Landsat: https://www.youtube.com/watch?v=Ezn1ne2Fj6Y

Citizen Scientists: Very sophisticated sensor; Play a key role in sustainability applications

Sensing Biodiversity, Science 346, 301 (2014); Woody Turner
The Citizen Science project at the Lab of Ornithology at Cornell empowers everyone interested in birds to contribute to research by submitting bird observations to the eBird webportal.
Citizen Science for Bird Conservation

- 300K+ volunteer birders
- 300M+ bird observations
- 22M+ hours of field work (2500+years)

eBird

CORNELL LAB of ORNITHOLOGY
Bird Distributions
Machine Learning and Citizen Science

State of the Birds Report
(officially released by Secretary of Interior)

Novel Approaches To Conservation Based on eBird Models

Distribution Models for 400+ species with weekly estimates at fine spatial resolution (3km²)

Adaptive Spatio-Temporal Machine Learning Models and Algorithms

Relate environmental predictors to observed patterns of occurrences and absences

300K+ volunteer birders
300M+ bird observations
22M+ hours of field work (2500+years)

Land Cover
Weather
Remote Sensing

Hemisphere Scale Bird Distribution Models, Revealing, at a fine resolution, Species’ Habitat Preferences

1st Time

Patterns of occurrence of the Barn Swallow for different months of the year Source: Daniel Fink
Dynamic High-Precision Conservation
Bird Returns

Protecting Migratory WaterBirds in California Against Drought by paying farmers to keep water in rice fields during the bird migration.

1. Pacific Migration Flyway
2. eBird Models
3. Target Estimates
4. Reverse Auction Bid Selection

Farmers submit bids to keep the rice fields flooded during short periods of bird migration in California.

Sacramento Valley, CA

Using big data, machine learning techniques, and a market-based approach, The Nature Conservancy has generated over 20,000 acres of additional habitat for waterbirds in California.

Radically novel, dynamic, way of doing bird conservation. Only possible because of advanced computational methods.
Data Bias Problem

Prevalent problem in citizen science
Collected data are often aligned with the participants’ preferences rather than scientific objectives.

How to incentivize Citizens to visit under-sampled areas?

Principal-Agent Framework

Pricing Problem: Bi-level Optimization Problem
Organizer has to learn agents’ preferences and constraints (to set the right incentives for the desired behavior)

Identification Problem: Agent directly sees Organizer’s prices or incentives

Avicaching: A Two Stage Game for Bias Reduction in Citizen Science

Field: Pilot Program

Incentivize eBirders to visit undersampled locations.

Incentives:
- Avicaching points,
- leaderboards
- Lotteries (e.g. binoculars.)
Socio-ecological corridor for Andean Bear in Ecuador

Planning Conservation Decisions in Ecuador
Integrating Socio-Economic Factors

Ecuador: Top Biodiversity Hotspots in the World

Challenge: SocioBioLinks Corridor Design

→ Protecting biodiversity
Population of Andean bears and other species

→ Improving the lives of local communities
(level of welfare/economic stability of local residents

→ Ecotourism, cooperatives, other activities

Andean Bear: Umbrella species
Habitat needs represent minimal requirements for numerous co-occurring species

Photo: Santiago Molina
Examples in Other Interdisciplinary Projects
Poverty Mitigation in Africa: Big Data for Africa

Improving Weather, Rangeland, and Forage Maps in Africa to protect farmers and herders

Deployment of 20,000 Low-cost Weather-Stations in African Schools

Herders Submit Vegetation Images and Surveys with Smartphones

Africa is very poorly sensed
(limited environmental data, vegetation maps, only a few reliable weather stations)
Dynamic Energy Storage and Economic Dispatch in the Presence of Renewables for the SMART Grid

Renewables are non-dispatchable

→ Modeling uncertainty
→ Robust plans combining renewables and traditional power resources

Multiple timescales

Due to, e.g.
→ Different types of generation
→ Fine grained frequency regulations (every few seconds),
→ Energy shifting (spanning hours) → Peak capacity charges (spanning year).
Accelerating Materials Discovery for Sustainable Energy
Accelerating the Development of Solar Fuel Generators

“... existing energy approaches [...] will not be enough to secure our energy future. Meeting the challenge will require [...] scientific breakthroughs in new materials....”


Solar fuel Cell

Solar fuels can be substantially produced and stored

Renewable transportation fuels can be synthesized using solar energy, water and carbon dioxide

Picture by John Gregoire

Caltech

Cornell High-Energy Synchrotron

JCAP

Joint Center for Artificial Photosynthesis

CHESS
Discovering better catalysts for solar fuel generators

Simultaneous synthesis of thousands of materials

Co-sputtering (similar to atomic spray painting)

Silicon Wafer

Metal1

Metal2

Metal3

High-Throughput Materials Discovery

How to determine the crystal structure of the materials, based on the X-ray diffraction patterns?

Crystal Structure Map

Problem: Infer the crystal structures of the materials from the X-ray diffraction patterns (currently a largely manual process, requiring expert knowledge)

Goal: Achieve High-Throughput Crystal Structure Identification

Difficulty: Often X-ray diffraction patterns correspond to a mixture of crystal structures

Challenging to un-mix the X-ray diffraction patterns
Can we use Citizen Science and Crowdsourcing to speed up the process of finding solutions?
Human Computation Framework for Boosting Computational Solvers

Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14
Decomposition: Slices of Sample Points
Different Visualizations of X-Ray Diffraction Patterns
Human Computation Framework for Boosting Combinatorial Solvers

Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14
Human Intelligent Task (HIT)

A worker is asked to identify patterns of similar vertical lines that intersect with sample 4 (whose detected intensity peaks are marked with red dots).

Complete Task

Orange pattern - 3 vertical lines.
Purple Pattern - 3 vertical lines.

Unmarked - all others are less clear and because of their ambiguity have been left unmarked, which is the correct decision, since workers are told to be conservative.
Human Computation Framework for Boosting Combinatorial Solvers

Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14
Human Computation Framework for Boosting Combinatorial Solvers
Human Computation Framework for Boosting Combinatorial Solvers

Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14
Human input can dramatically speed up the performance of combinatorial optimization methods.

Approaches leverages the complementary strength of human input, providing global insights into problem structure, and the power of combinatorial solvers to exploit complex constraints.

<table>
<thead>
<tr>
<th>System</th>
<th>P</th>
<th>$L^*$</th>
<th>K</th>
<th>#var</th>
<th>#cst</th>
<th>Time w/o user input (s)</th>
<th>Time w/ user input (s)</th>
<th># assigned var. by user</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/B/C</td>
<td>36</td>
<td>8</td>
<td>4</td>
<td>408</td>
<td>2095</td>
<td>3502</td>
<td>859</td>
<td>19 (4.6%)</td>
</tr>
<tr>
<td>A/B/C</td>
<td>60</td>
<td>8</td>
<td>4</td>
<td>624</td>
<td>3369</td>
<td>17345</td>
<td>4377</td>
<td>18 (2.9%)</td>
</tr>
<tr>
<td>Al/Li/Fe</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>267</td>
<td>1009</td>
<td>79</td>
<td>4</td>
<td>6 (2.2%)</td>
</tr>
<tr>
<td>Al/Li/Fe</td>
<td>28</td>
<td>6</td>
<td>6</td>
<td>436</td>
<td>1864</td>
<td>346</td>
<td>62</td>
<td>12 (2.7%)</td>
</tr>
<tr>
<td>Al/Li/Fe</td>
<td>28</td>
<td>8</td>
<td>6</td>
<td>490</td>
<td>2131</td>
<td>100</td>
<td>1003</td>
<td>30 (4.2%)</td>
</tr>
<tr>
<td>Al/Li/Fe</td>
<td>28</td>
<td>10</td>
<td>6</td>
<td>526</td>
<td>2309</td>
<td>281</td>
<td>1003</td>
<td>30 (4.2%)</td>
</tr>
<tr>
<td>Al/Li/Fe</td>
<td>45</td>
<td>7</td>
<td>6</td>
<td>693</td>
<td>3281</td>
<td>188</td>
<td>1003</td>
<td>30 (4.2%)</td>
</tr>
</tbody>
</table>

Table 1: Runtime of SMT solver with and without user input. $P$ is the number of sample points, $L^*$ the average number of peaks per phase, $K$ the number of basis patterns, #var the number of variables and #cst the number of constraints.

Human Computation: From ~13 hours to 1000 seconds!
Real System: Pd-Ta-Rh system

198 synchrotron x-ray diffraction patterns
Task #116 (1/17)

Research Participant Disclosure: This task is part of a study being conducted by Cornell University researchers and data will be collected for research purposes.

Identify Patterns of Vertical Lines

Identify the most prominent pattern (Pattern A) and, if applicable, up to two additional patterns (Pattern B and Pattern C) intersecting the Target Row. Mark the most definite and identifiable vertical lines/blobs by clicking on them with the appropriate mode selected.

Instructions (click to show/hide)

Example (click to show/hide)

Definitely in Pattern A
Definitely in Pattern B
Definitely in Pattern C

Messages:

Any comments or notes on this particular image?

Any feedback on the instructions or the design of the task?
Participants ignored noisy peaks; Large number aggregated the peaks into single pattern (8 or more peaks)
Pd-Rh-Ta System: Best solution: Hybrid Human-Machine System

Ortho-$(\text{Pd}, \text{Rh})_2 \text{Ta}$

$\text{Pd}_{0.14} \text{Rh}_{0.40} \text{Ta}_{0.46}$ → best known Pt-free catalyst for methanol oxidation
Agriculture

Nitrogen Based Fertilizers
Nitrogen Based Fertilizers

Dead Zone in Gulf of Mexico

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

Study of fertilizers and design of Experiments

Collaborator Harold van Es
To Put It Into Perspective with Rounded Numbers….

- Average annual N$_2$O losses per acre (corn lands)- 7.5 lbs
  Equivalent to
  – combustion of 126 gallons of gasoline
  – 3,700 miles of driving an average passenger car

- Assuming a farm with 500 ac of corn, the annual global warming impact is equivalent to about 1.8 million miles of driving (70 times around earth).

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

- 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).

Spatially Balanced Latin Square
Diagonally Symmetric SBLS

Observed symmetric SBLS of Order 6

Imposing symmetry could generate SBLS of order 8 and 9
Streamlining Constraint Reasoning

Discovery of structural properties across solutions.

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

Can we find an domain independent way of streamlining?

YES: XOR-Streamlining based on random parity constraints
(Valiant and Vazirani 1986, Unique SAT)

Counting, Sampling, Inference.

Gomes, Sabharwal, Selman, 2006
Ermon, Gomes, Sabharwal, Selman, 2013, 2104, 2015
### Discovery of Construction for Spatially Balanced Latin Squares: Streamlining Reasoning + Human Computation

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>14</td>
<td>10</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>14</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>11</td>
<td>13</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>11</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>13</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>13</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>12</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>5</td>
<td>3</td>
<td>11</td>
<td>10</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>7</td>
<td>1</td>
<td>9</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>2</td>
<td>7</td>
<td>13</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>6</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>1</td>
<td>11</td>
<td>8</td>
<td>2</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>8</td>
<td>10</td>
<td>1</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>13</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>2</td>
<td>14</td>
<td>3</td>
<td>9</td>
<td>8</td>
<td>4</td>
<td>13</td>
<td>1</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>12</td>
<td>1</td>
<td>14</td>
<td>2</td>
<td>11</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>12</td>
<td>3</td>
<td>11</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

```
for row $i = 1, \ldots, N$ do
    $k = 1$;
    $j = 1$;
    $a_{i,j} = i$;
    while $j < N$ do
        if $k$ is odd then
            while $a_{i,j} + i \leq N$ and $j < N$ do
                $a_{i,j+1} = a_{i,j} + i$;
                $j = j + 1$;
            else
                while $a_{i,j} - i \geq 1$ and $j < N$ do
                    $a_{i,j+1} = a_{i,j} - i$;
                    $j = j + 1$;
            if $j < N$ then
                if $k$ is odd then
                    $a_{i,j+1} = 2N + 1 - i - a_{i,j}$;
                else
                    $a_{i,j+1} = i - a_{i,j}$;
                    $k = k + 1$;
                    $j = j + 1$;
        Algorithm: SBLS-sequence procedure for SBLS of order $N$, when $2N + 1$ is prime.
```
Tools for Site-Specific Adaptive Nitrogen Management

Two-Way Street: Smartphones & Tablets and the Cloud

Adapt N - Infrastructure

High Resolution Climate Data:
- Currently available for Northeast & Iowa
- Expanding to the Midwest
- Enables Field-Scale adaptation

Metereologic data

Harold van Es
CompSustNet: Cross-cutting Computational Themes
CompSustNet: 3 Core Sustainability Themes
Over 20 Novel Interdisciplinary Research Projects (IRPs)

I Conservation and Biodiversity

Prioritizing Conservation Decisions
Protecting Endangered Species

II Balancing Socio-economic Needs

Protecting Farmers in Africa by Improving Weather System
Protecting Migratory Herders in Africa

III Renewable Energy and sustainable materials

Accelerating the Discovery of New Sustainable Energy Materials
Energy Storage and Integration of Renewables In Smart Grid

Wide range of sustainability applications:
To build a *coherent and cohesive computer science sub-field* it’s paramount to
*go deep* in terms of *Cross-Cutting Core Computational Problems* and
apply them to different sustainability areas.
Cross-Cutting Computational Theme: Pattern Decomposition in Big Data

High Throughput Materials Structure Identification

Dimensionality reduction with complex constraints

Similar computational problems:
Both involve the decomposition of a signal (X-rays / audio) into patterns (crystal structures / types of sounds)

Our Philosophy:
Students should work on similar computational problems in different domains.
Grad students working on materials discovery also work on the elephant calls.

Elephant Listening Project
Project to monitor elephant populations by analyzing forest recordings

We represent these cross-cutting computational themes with colored “subway lines.”
Pattern Decomposition in Big Data is the blue line.
Cross-Cutting Computational Theme

Findings can be transferred across domains!

Mechanism Design: Principal-Agent Game-Theory Model

Green Security Games: Preventing Poaching and Illegal Fishing

How to incentivize Citizens to visit under-sampled areas?

Computationally, these problems have similarities.

Where to place patrols to prevent poaching and illegal fishing?

We represent these cross-cutting computational themes with colored “subway lines.”

Mechanism design is the baby blue line.
Conclusions
1st NSF Expeditions in Computational Sustainability

Expedition allowed us:

- To nucleate the Computational Sustainability field
- To identify a number of core research directions for maximal impact, both in terms of Computer Science and Sustainability.

Big Challenge:
How to get involved and navigate in this highly interdisciplinary area.

Growing number of computer scientists eager to get involved in computational sustainability research, but don’t have the connections and access to sustainability projects.
CompSustNet: Large Research Network for Expanding Horizons of Computational sustainability

**Research**
- Optimizing Dynamical Models Simulation
- Big Data Machine Learning
- Crowdsourcing Citizen Science

**Research Community Building**
- Web Portal
- Catalog Of Problems
- Annual Conference
- Host visiting Scientists

**CompSust.Net**
- Coordinating transformative synthesis collaborations
- Interdisciplinary Research Projects (IRPs)

**Education**
- Postdocs
- Honors projects
- Research seminar series

**Outreach**
- Citizen Science Projects
- K-12 Outreach
- Diversity

**General Public Outreach**
- Engaging Gov., NGOs Institutions and Companies
- Broad Dissemination of scientific results

**Diversity**
- Research seminar series
- Students’ conference

**Doctoral students**
- Summer REU program targeting minority students
- Distributed Courses
Annual CompSust Conference
Bringing together students, researchers and practitioners

Doctoral Program and Tutorials

CompSustNet Web Portal
• Catalog of core research problems
• Datasets and open source software
• Publications
• Wiki for collaboration
• Course resources
• Discussion & e-mail list

UDiscoverIt
• Citizen science participation
• Crowdsourcing tasks

At major conferences:
Refereed CompSust Tracks, DataSet papers, Workshops, Doctoral programs, Panels and tutorials;

CompSust-2016
4th International Conference on Computational Sustainability
July 6-8, 2016
Cornell University, Ithaca, NY
Computational Sustainability:
High potential to advance state-of-the-art of Computer Science
Driven by concrete sustainability challenges

1. New challenging problems
2. New formalisms and concepts from other disciplines

→ New Core Paradigmatic problems in Comp. Sci.

→ Broadening the Scope & Diversity in CS

Computational Thinking that will provide new insights, methodologies, and solutions to sustainability problems

→ Societal Impact

Thank you!