WELCOME TO TACC

REU Community Coding 2023

Introductions - Icebreaker

Charlie Dey (*TACC*)

Je'aime Powell (*TACC*)

TACC AT A GLANCE

- Research center located at UT Austin
- ~175 Staff (~70 PhD scientists, ~20 students)
- Funded by UT System, NSF (85% external grants)

- **Users:** >10,000 on 2,300 active projects across all fields
- **Partnerships:** UT Research Cyberinfrastructure (UTRC), Extreme Science and Engineering Discovery Environment (XSEDE), Industry, International

Mission: "To enable discoveries that advance science and society through the application of advanced computing technologies."

TACC AT A GLANCE

- **Capacity and Infrastructure:**
  - A billion compute hours per year
  - >5 billion files, >50 petabytes of data
  - Hundreds of public datasets
  - 18 MW data center

- **Systems and Services:**
  - High performance computing (HPC), high throughput computing (HTC), large scale data storage, cloud computing, visualization
  - Portals and gateways, web service APIs, rich software stacks
  - Consulting, curation and analysis, code optimization, training and outreach

AN ECOSYSTEM FOR EXTREME SCALE SUPERCOMPUTING

Stampede-2
#51 HPC system,18PF, 350k cores

Lonestar 6
Texas-focused HPC/HTC
71k AMD Epyc 7763 cores, 5TF, 20+ GPUs

Hikari (RIP)
Protected Data Containers
10,000 Intel Haswell cores 400TF

Wrangler
Data Intensive Computing
0.6 PB flash storage 1 TB/s read rate

Maverick2
GPU/Interactive/Analytics
GeForce GPUs, Jupyter and interactive support

Jetstream
w/ Indiana U.
Science Cloud/HTC
VM Library
~10,000 Intel Haswell cores

Stockyard
Shared Storage Across TACC
40PB, Lustre

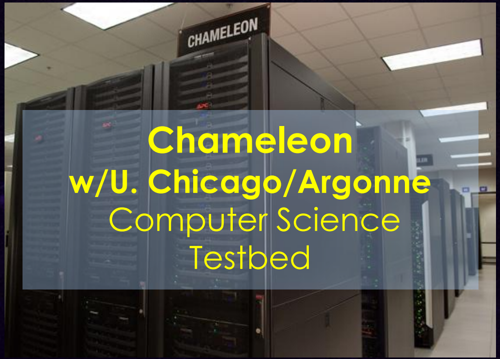
Ranch
Archive
HIPAA-Aligned
30PB Disk Cache,
0.5EB Tape

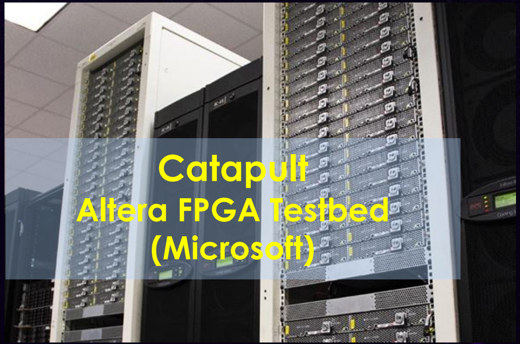
Corral
Published Data Collections
HIPAA-Aligned
20PB Replicated Disk,

Stallion

Lasso

EXPERIMENTAL SYSTEMS





"COMPUTING FOR THE ENDLESS FRONTIER"

Dan Stanzione

Executive Director, Texas Advanced Computing Center

Associate Vice President for Research, The University of Texas at Austin

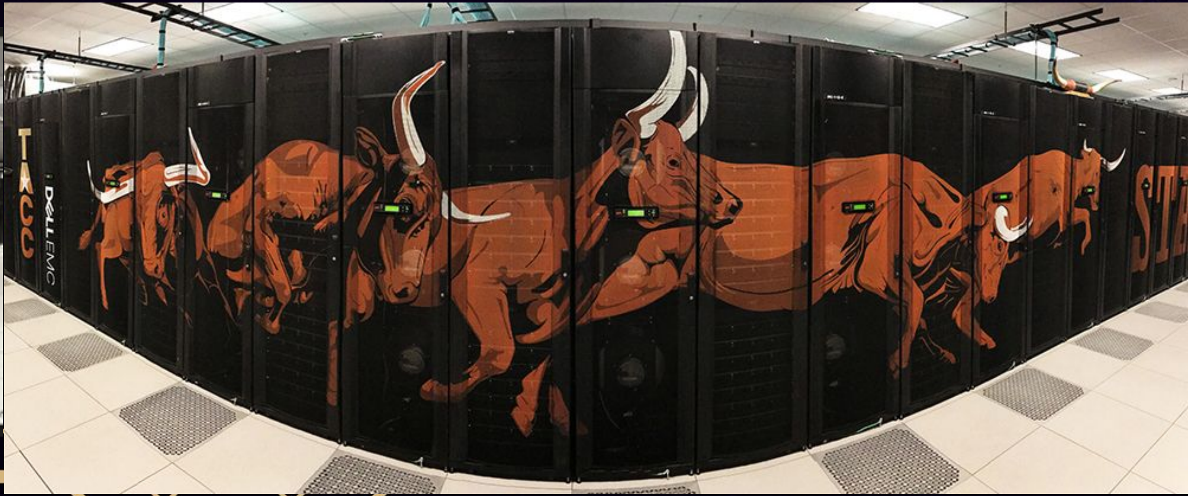

FRONTERA
TACC
NSF
TEXAS

RAPID GROWTH FROM THEN TO NOW…

- 2003 – First Terascale Linux cluster for open science (#26)
- 2004 – NSF funding to join the Teragrid
- 2006 – UT System Partnership to provide Lonestar-3 (#12)
- **2007 - $59M NSF award – largest in UT history – to deploy Ranger, the world's largest open system (#4)**
- 2008 – funding for new Vis software and launch of revamped visualization lab.
- 2009 - $50M iPlant Collaborative award (largest NSF bioinformatics award) moves a major component to TACC, life sciences group launched.
  - In 2009, we reached, 65 employees.

NOW, A WORLD LEADER IN CYBERINFRASTRUCTURE

- 2010, TACC becomes a core partner (1 of 4) in XSEDE, the TeraGrid Replacement
- 2012, Stampede replaces Ranger with new $51.5M NSF Award
- 2013, iPlant is renewed, expanded to $100M
- 2015, Wrangler, first data intensive supercomputer is deployed.
- 2015, Chameleon cloud is launched
- 2015, DesignSafe, the cyberinfrastructure for natural hazard engineering, is launched.
- 2016 Stampede-2 awarded the largest academic system in the United States, 2017-2021.



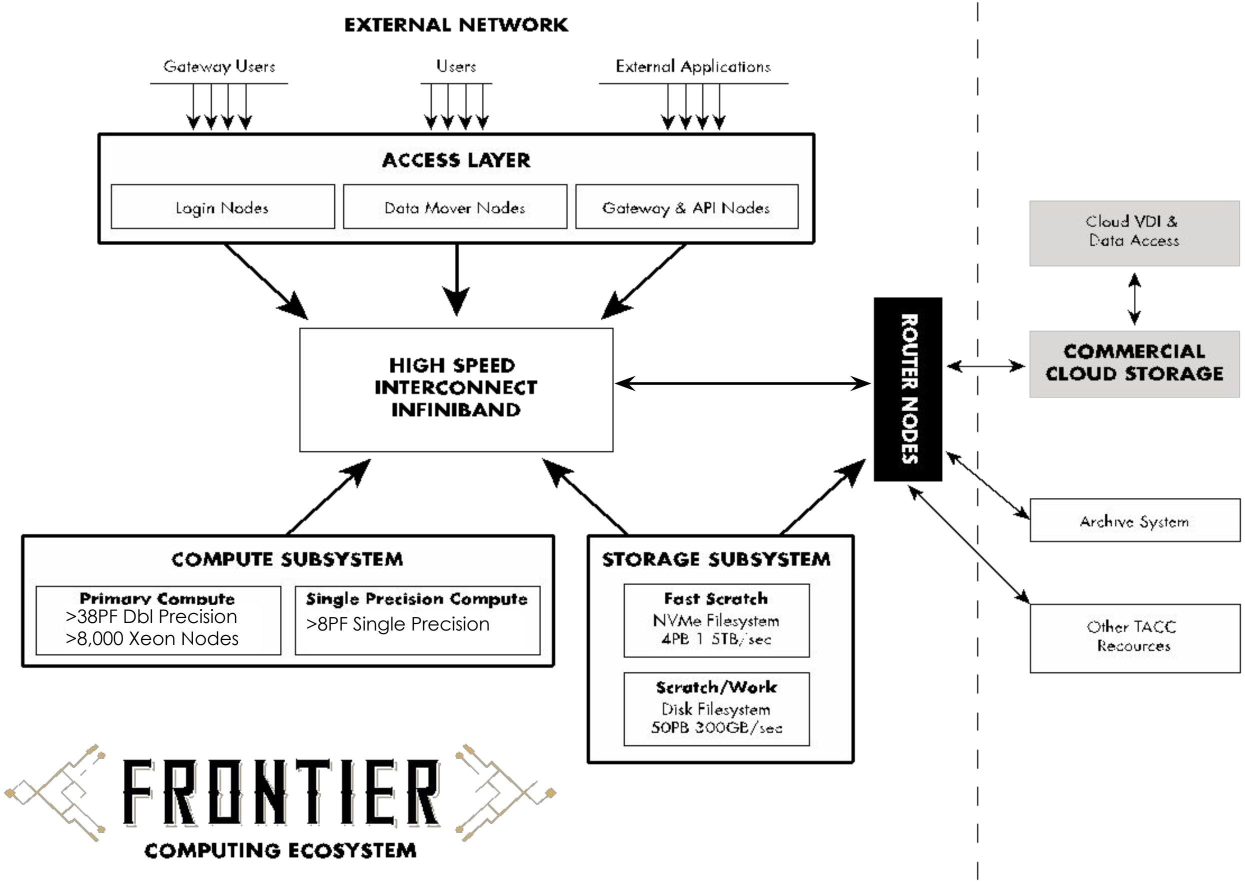
FRONTERA SYSTEM --- PROJECT

- Deploy a system in 2019 for the largest problems scientists and engineers currently face.
- Support and operate this system for 5 years.
- Plan a potential phase 2 system, with 10x the capabilities, for the future challenges scientists will face.

FRONTERA SYSTEM --- HARDWARE

- 8th Fastest Supercomputer in the world
  - #1 for Open Science
- Primary compute system: DellEMC and Intel
  - 35-40 PetaFlops Peak Performance (Next Generation Xeon processors)
- Interconnect: Mellanox HDR and HDR-100 links.
  - Fat Tree topology, 200Gb/s links between switches.
- Storage: DataDirect Networks
  - 50+ PB disk, 3PB of Flash, 1.5TB/sec peak I/O rate.
- Single Precision Compute Subsystem: Nvidia
- Front end for data movers, workflow, API

SYSTEM OVERVIEW



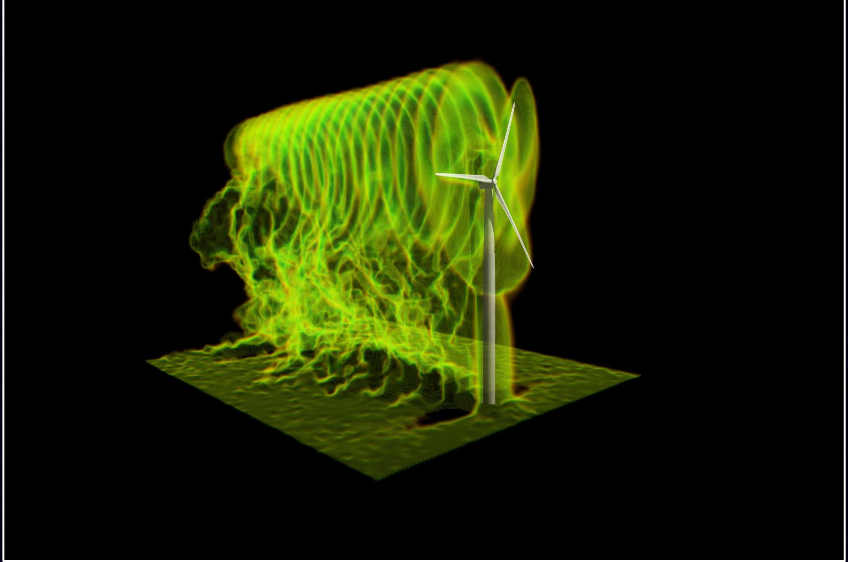
- **Humphry Davy, Inventor of Electrochemistry, 1812**
- (Pretty sure he was talking about our machine).

> Nothing tends so much to the advancement of knowledge as the application of a new instrument. The native intellectual powers of men in different times are not so much the causes of the different success of their labours, as the peculiar nature of the means and artificial resources in their possession.
>
> Humphry Davy

PICTUREQUOTES.com

REAPING POWER FROM WIND FARMS

Multi-Scale Model of Wind Turbines



- Optimized control algorithm improves design choices
- New high-res models add nacelle and tower effects
- Blind comparisons to wind tunnel data demonstrate dramatic improvements in accuracy
- Potential to increase power by 6-7% ($600m/yr nationwide)

"TACC...give[s] us a competitive advantage…"

Graphic from Wind Energy, 2017.

Christian Santoni, Kenneth Carrasquillo, Isnardo Arenas-Navarro, and Stefano Leonardi

UT Dallas, US/European collaboration (UTRC, NSF-PIRE 1243482)

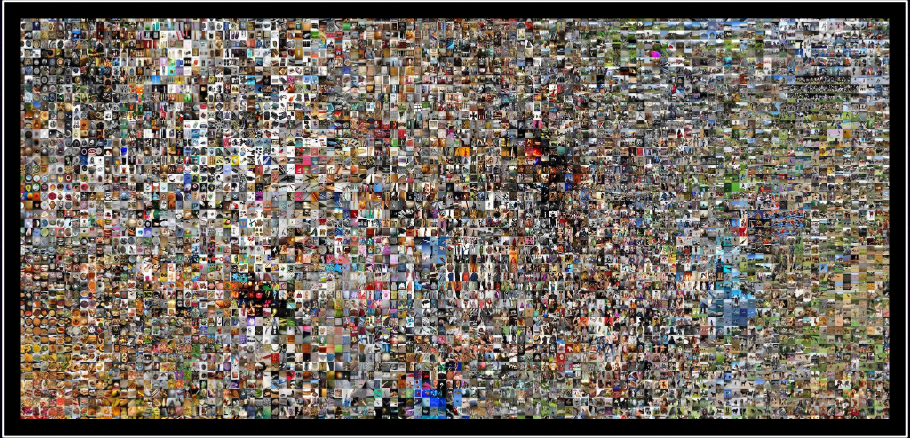
TACC Press Release

"Using commodity HPC servers...the time to data-driven discovery is reduced and overall efficiency can be significantly increased." (Niall Gaffney, TACC)

RECORD ACHIEVED ON AI BENCHMARK

TACC, Berkeley, Cal Davis collaborate on large-scale AI runs

Graphic credit Andrej Karpathy



- Research demonstrating the potential of commodity hardware for AI
- Skylake ImageNet benchmark: (100 epochs, 11 min, 1024 nodes) -- fastest result at time of publication
- Knights Landing ImageNet benchmark (90 epochs, 20 min, 2048 nodes) – 3x faster than Facebook, with higher large-batch accuracy

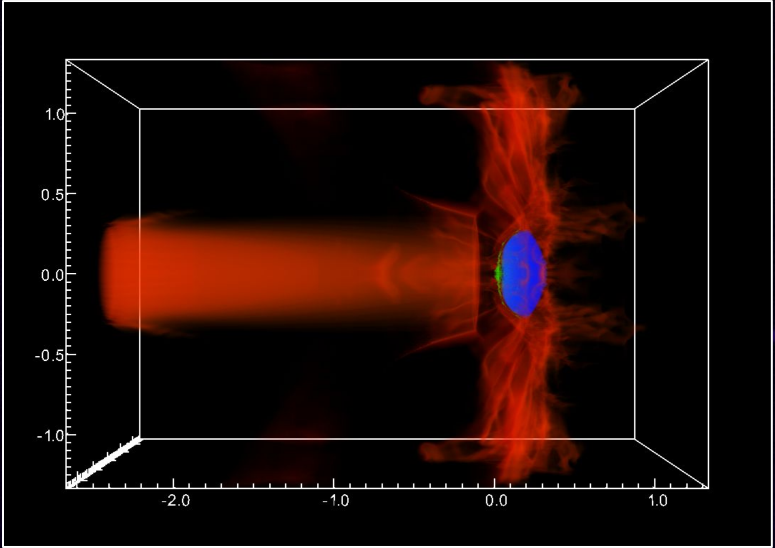
Yang You, Zhao Zhang, Cho-Jui Hsieh, James Demmel, Kurt Keutzer

TACC Press Release

USING KNL TO PROBE SPACE ODDITIES

Ongoing XSEDE collaboration focusing on KNL performance for new, high-resolution version of COSMOS MHD code

- Vectorization and other serial optimizations improved KNL performance by 50%
- COSMOS currently running 60% faster on KNL than Stampede1
- Work on OpenMP-MPI hybrid optimizations now underway
- Impact of performance improvements amounts to millions of core-hours saved



"The science that I do wouldn't be possible without resources like [Stampede2]...resources that certainly a small institution like mine could never support. The fact that we have these national-level resources enables a huge amount of science that just wouldn't get done otherwise." (Chris Fragile)

TACC Press Release

XSEDE ECSS: Collaboration between PI Chris Fragile (College of Charleston) and Damon McDougall (TACC)

MASSIVE DATA SET WORTHY OF ROSS ICE SHELF ITSELF

TACC partners with Lamont-Doherty Earth Observatory (LDEO) to host for one of the country's largest earth sciences data collections

- Managing hundreds of TB using Stampede2, Corral, and Ranch: storage, provenance, visualization, and public access
- Achieved 10x workflow speedup by moving to TACC (from 50 hrs down to 5 hrs for transfer and analysis tasks)

"...partnership...with TACC shows [it's] possible to manage…this level of data in a cost-effective, user-friendly and easily accessible manner…"

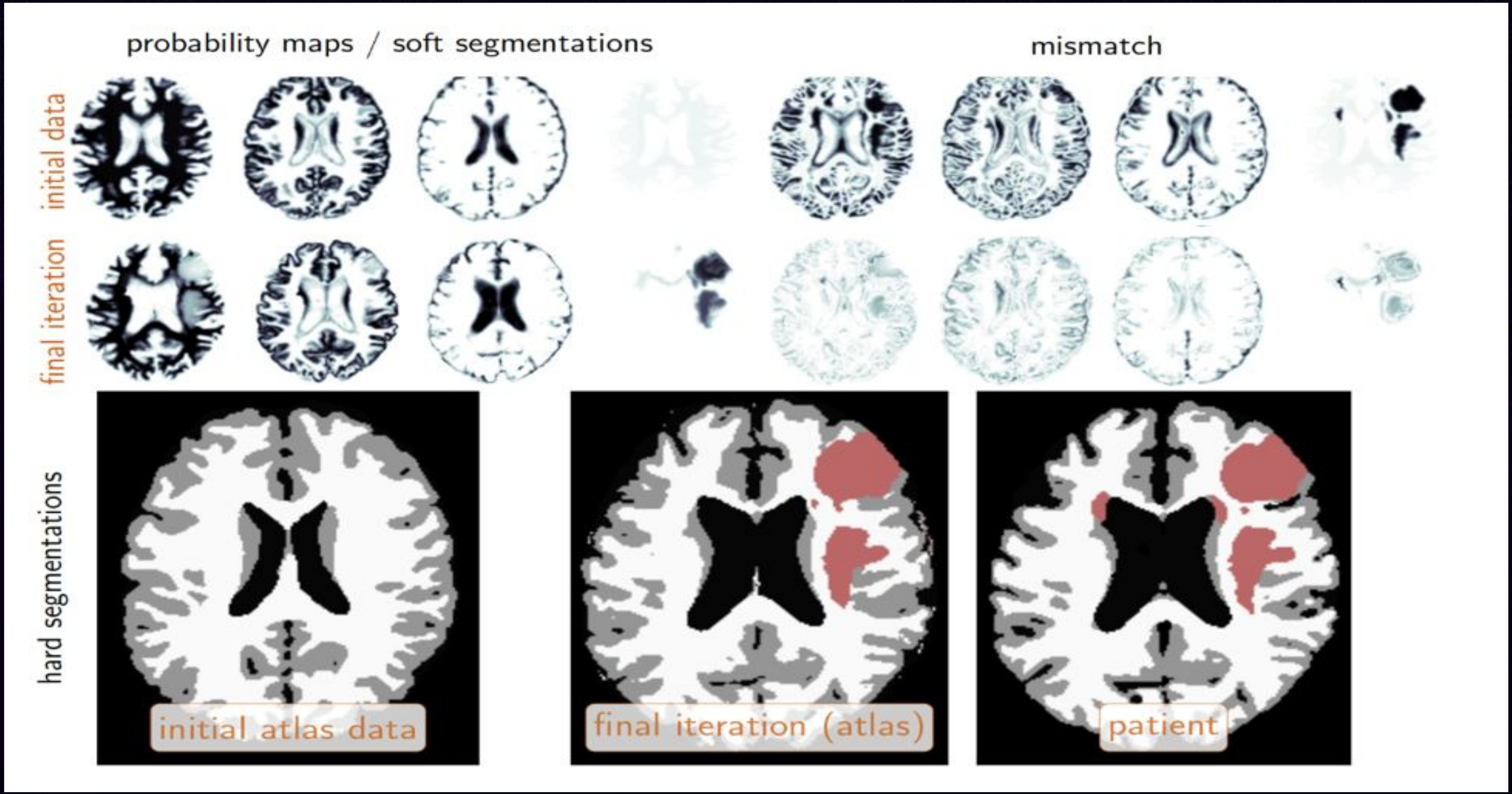
Image courtesy Oceanwide Expeditions.

TACC Press Release

PI Lingling Dong, Columbia University
XSEDE support to multidisciplinary, multi-institutional Rosetta project

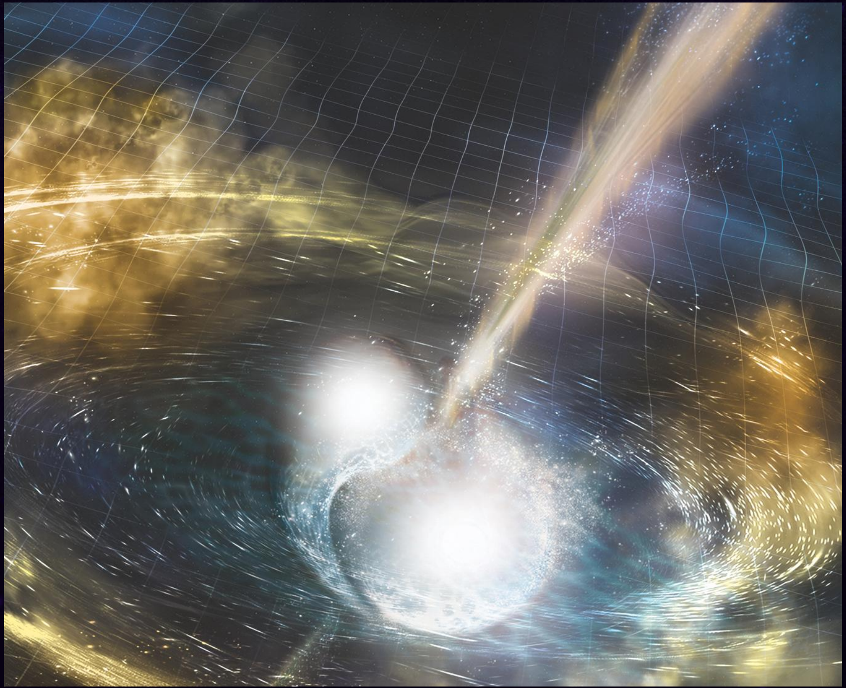
BRAIN TUMOR SEGMENTATION

- A team of researchers led by George Biros from The University of Texas at Austin scored in the top 25% of participants in the Multimodal Brain Tumor Segmentation Challenge 2017 (BRaTS'17) enabled by Stampede2 and other TACC resources.
- In the challenge, research groups presented methods and results of computer-aided identification and classification of brain tumors, as well as different types of cancerous regions.
- The team's method combined biophysical models of tumor growth with machine learning algorithms for the analysis of Magnetic Resonance imaging data of glioma patients.


probability maps / soft segmentations
mismatch
initial data
final iteration
hard segmentations
initial atlas data
final iteration (atlas)
patient

The Laser Interferometer Gravitational-Wave Observatory (LIGO)

waves.

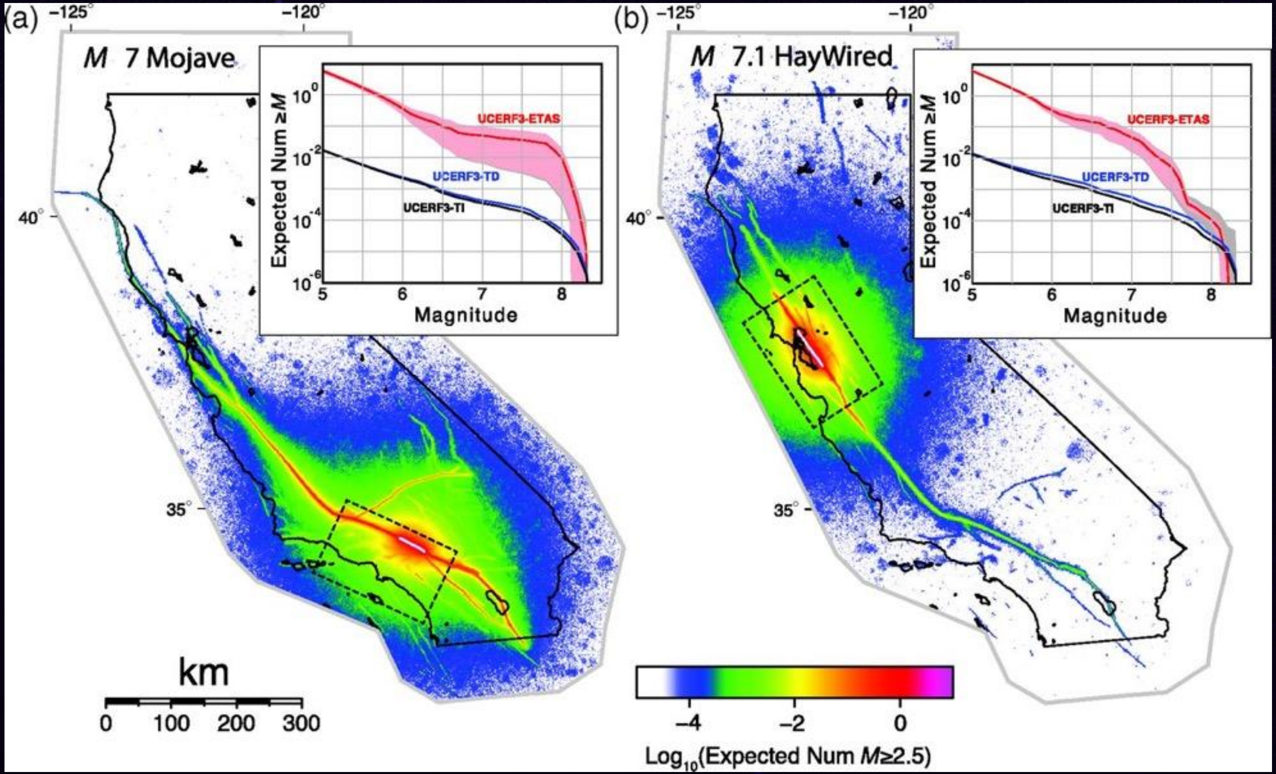


- TACC provided consulting services to LIGO researchers, 10x improvement in runtime
- seven million core hours on Stampede to help analyze the first detected gravitational

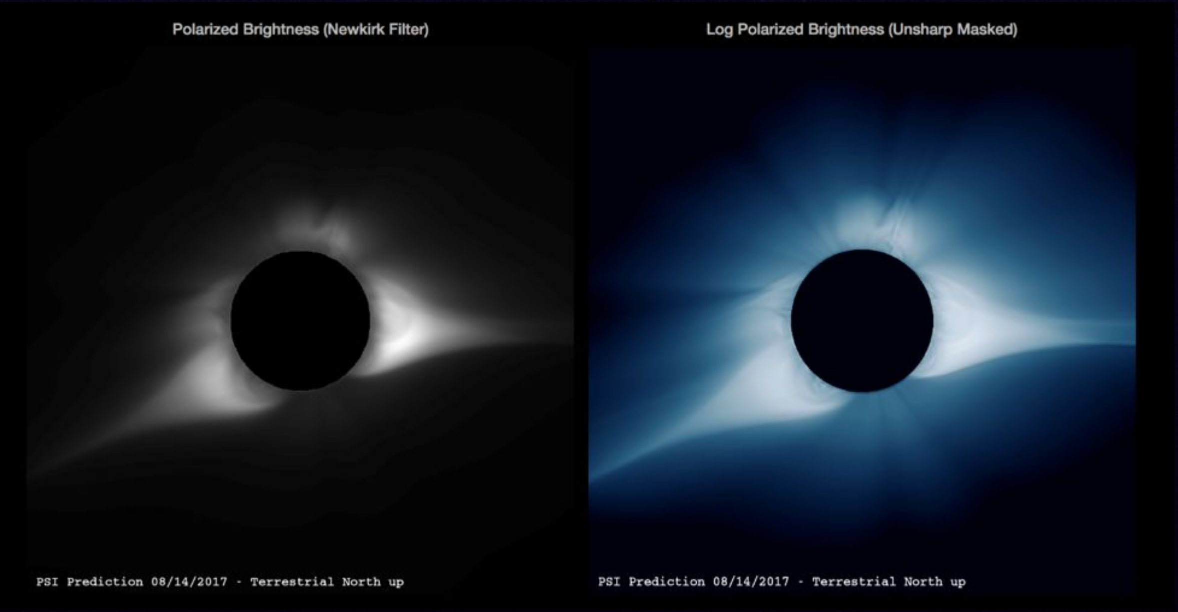
"Stampede was an excellent tool for improving our understanding of the universe we live in, from the smallest scale of sub-atomic particles to detecting gravitational waves that have traveled a million light-years to the earth, and a lot of exciting science and engineering in between"

Stuart Anderson, Caltech
LIGO Research Manager

AFTERSHOCK FORECASTING

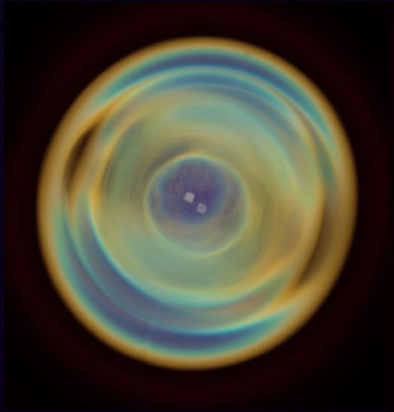


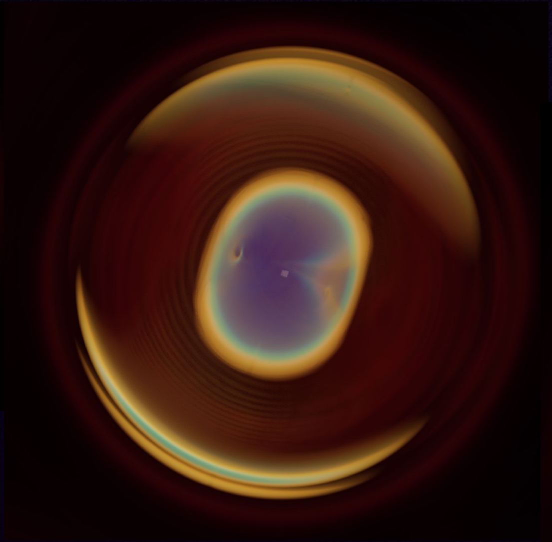
SOLAR CORONA PREDICTION



- Predictive Science, Inc. (California)
- Supporting NASA Solar Dynamics Observatory (SDO)
- Predicted solar corona on S2 during 8/21/17 eclipse

COSMOS GRAVITATIONAL WAVES STUDY





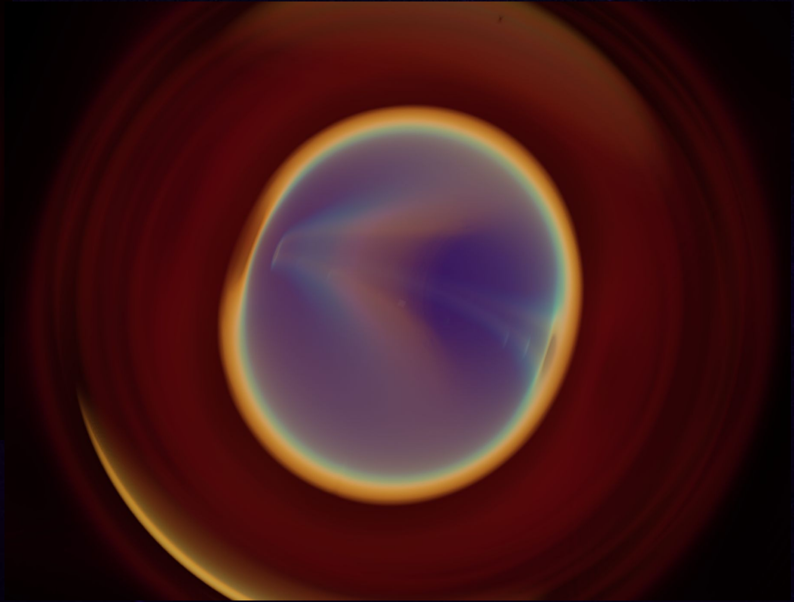


Image Credits:

Greg Abram – TACC

Francesca Samsel – CAT

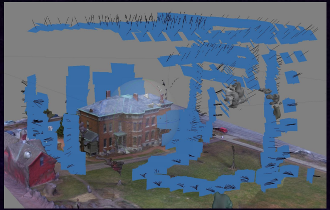
Carson Brownlee - Intel

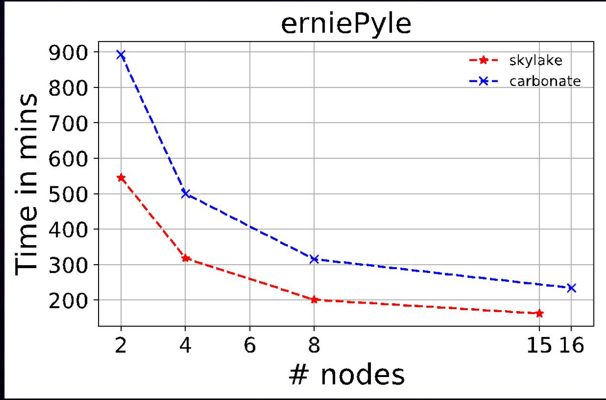
Markus Kunesch, Juha Jäykkä, Pau Figueras, Paul Shellard

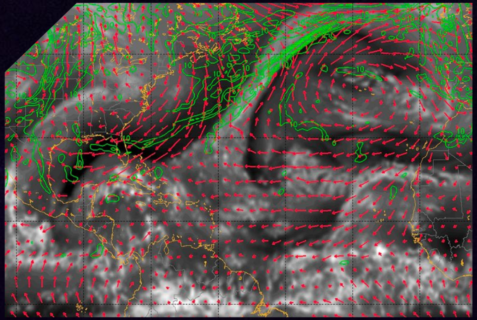
Center for Theoretical Cosmology, University of Cambridge

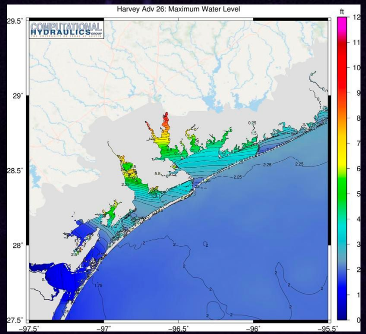
PHOTOGRAMETRY ON KNL

- Effort lead by IU (Wernert, McCombs, Ruan, Tuna)
- Create 3d point cloud & Mesh Model of texture/color map using tiled 2d images
  - Camera panoramas, Drone Survey
  - Future underwater shipwrecks/reefs
- Using Agisoft Photoscan software
  - More speedup from larger datasets
- Exploring OpenSource alternatives
  - Adding MPI layers needed



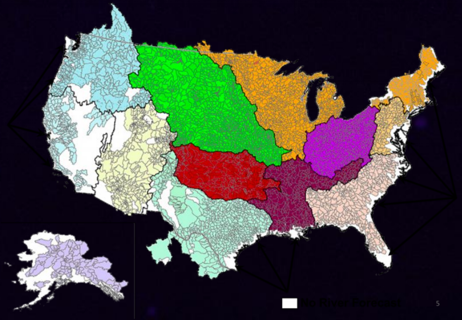


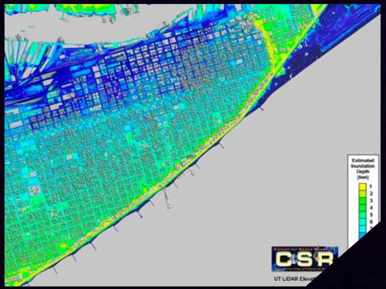




HARVEY

- Next Generation Storm Forecasting (with Penn State)
- Storm Surge Modeling (with Clint Dawson UT Austin)
- Preliminary river flooding and inundation maps (David Maidment UT Austin)
- Remote Image Integration and Assimilation (Center for Space Research, UT Austin)



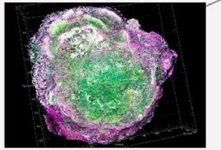


The Texas Advanced Computing Center accelerates basic and applied cancer research to help save lives.

fighting

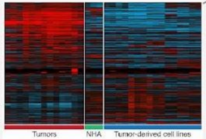
with supercomputers

Computer Modeling



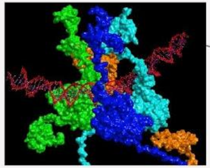
Researchers use advanced computing to model tissues, cells and drug interactions, and to design patient-specific treatments and identify new medicines.

Big Data Analysis



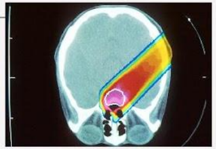
Supercomputers allow researchers to find patterns in genomes and among patient outcomes to pinpoint risks and target treatments.

Molecular Dynamics Simulations



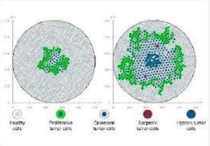
Simulating protein and drug interactions at the atomic level enables scientists to understand cancer and design more effective therapies.

Quantum Calculations



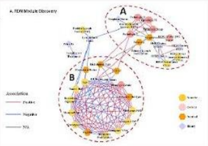
Exploring how proton and x-ray beams interact with DNA on the quantum level helps explain why radiation treatments work and how they can be optimized.

Trial Design



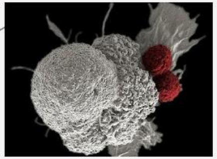
Researchers use TACC's advanced computers to design clinical trials that can determine the combination of dosages that will be most effective.

Clinical Planning



Supercomputers can test thousands of potential treatments in advance to help decide which one will work best.

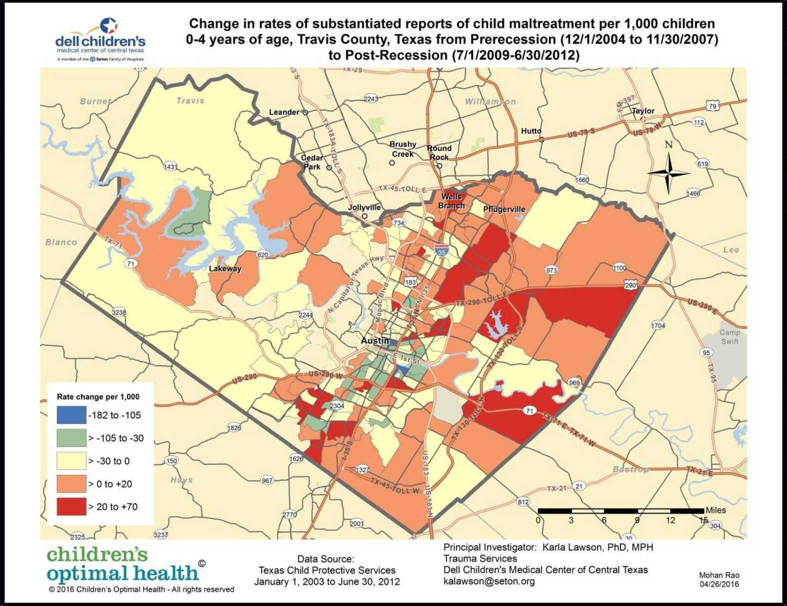
Artificial Intelligence



AI on high-performance computers can uncover relationships among complex cellular networks and reverse-engineer interventions.

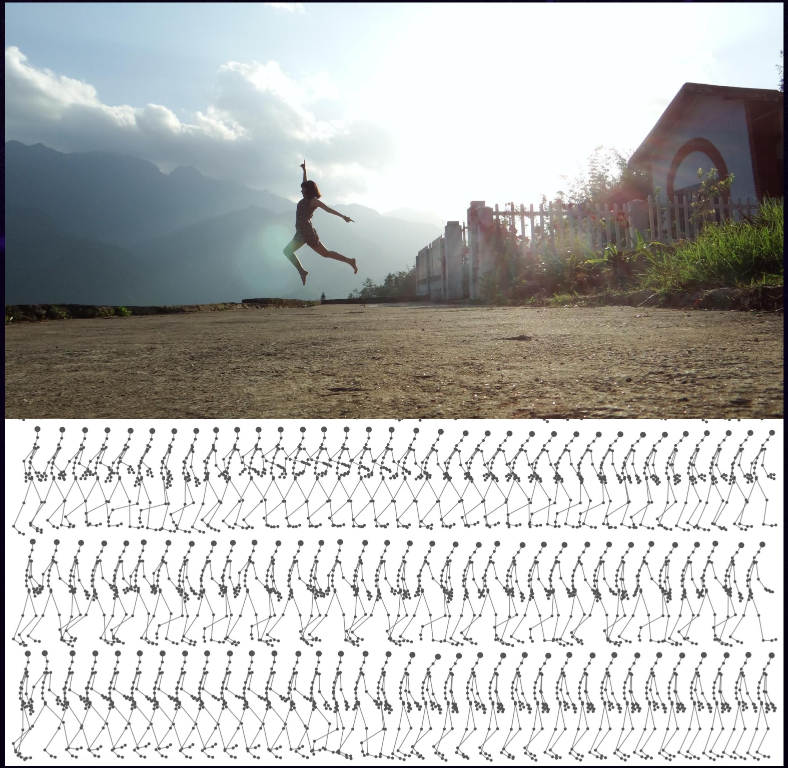
TEXAS ADVANCED COMPUTING CENTER

IMPROVE CHILD WELFARE THROUGH ADVANCED COMPUTING

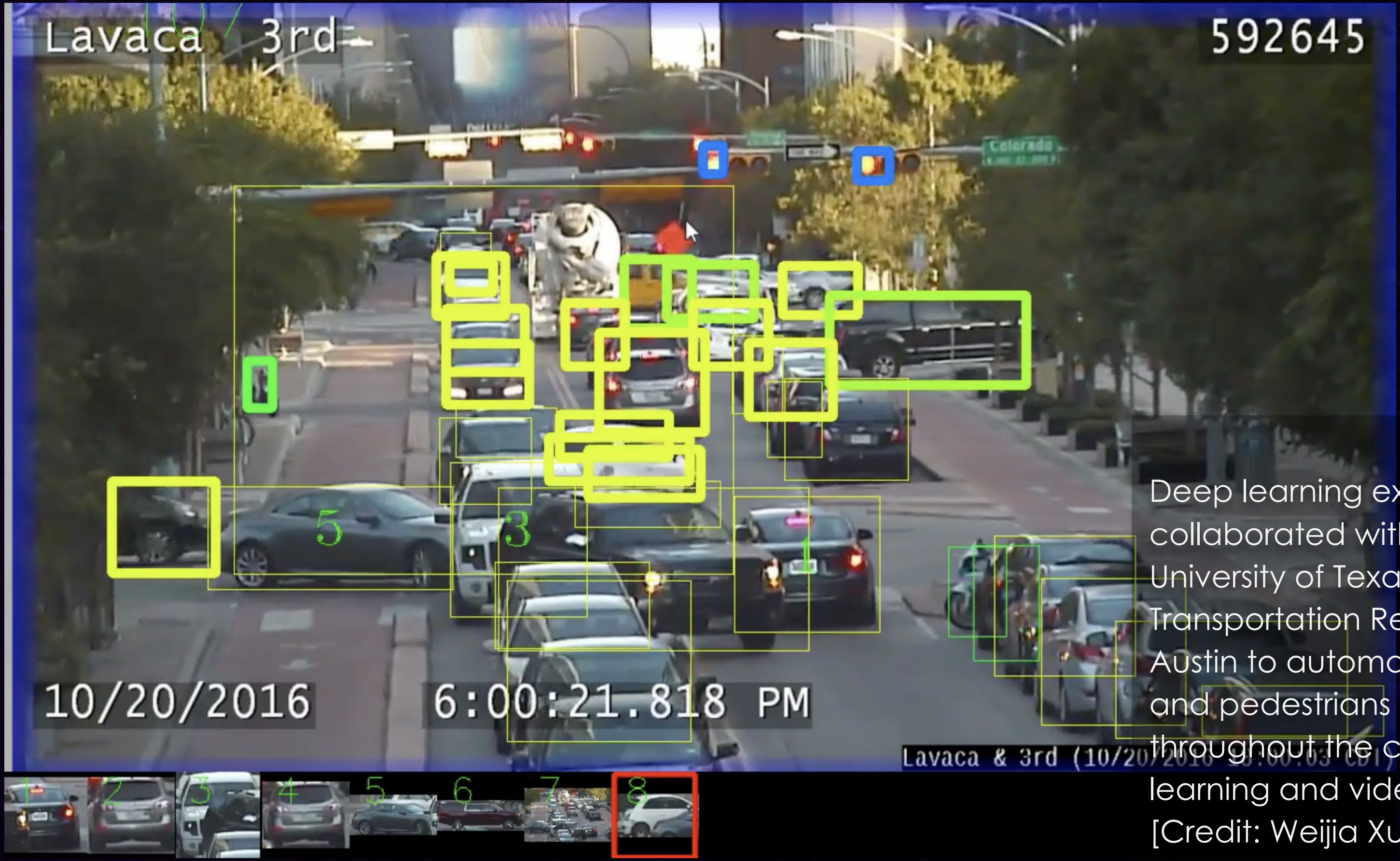


The Texas Advanced Computing Center (TACC) at The University of Texas at Austin has teamed up with Austin, Texas-based non-profit Children's Optimal Health (COH) to provide the technical infrastructure needed to help solve issues in children's health and education throughout Central Texas, including disease, mental health and adverse childhood experiences.

A DANCE WITH ALGORITHMS



- XSEDE resources help researchers merge art and science to create systems that can understand and produce human-quality movement
- Uses deep learning techniques and relies on the collaboration of graduate students with artists to create new algorithms
- *On a single computer, running our algorithms would take years, on medium-sized resources months, but using XSEDE, we can train some of most complex models within 24 hours. -* Philippe Pasquier, professor and researcher, Simon Fraser University

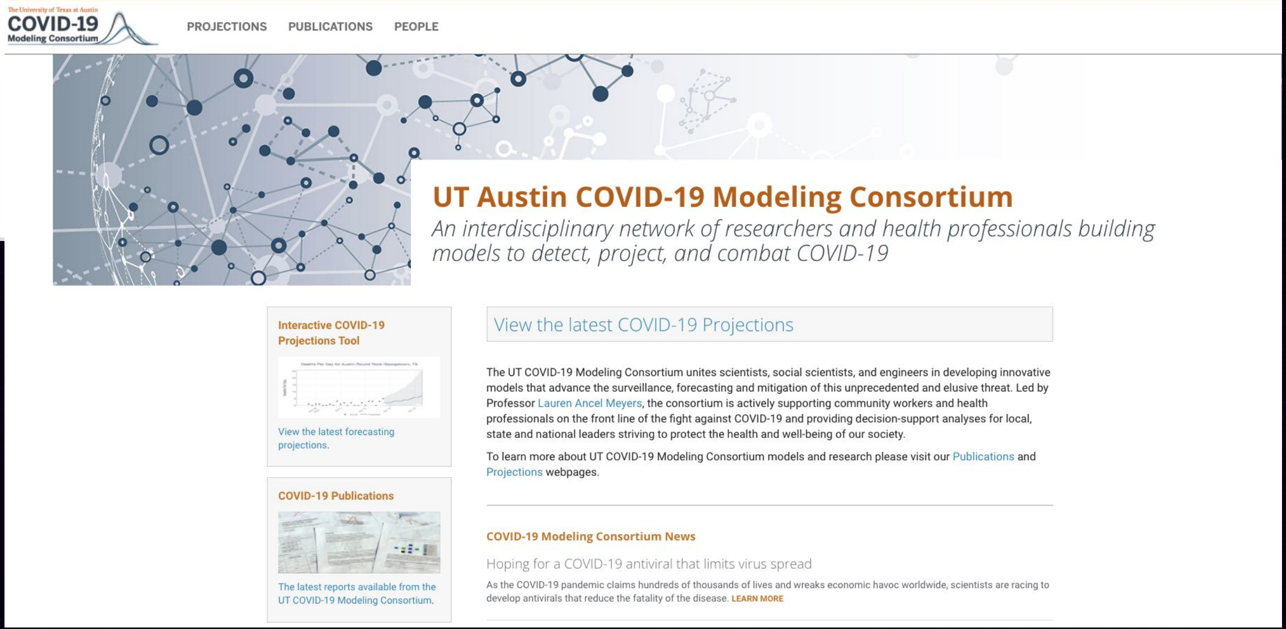


Deep learning experts from TACC collaborated with researchers at the University of Texas Center for Transportation Research and the City of Austin to automatically detect vehicles and pedestrians at critical intersections throughout the city using machine learning and video image analysis. [Credit: Weijia Xu, TACC]

Background: The COVID-19 pandemic required a rapid response to provide forecasting models to key decision makers. The COVID-19 Science Gateway is part of The University of Texas at Austin COVID-19 Modeling Consortium

Objective: Provides decision-makers with key predictive models necessary. Enables researchers to run models easily through the science gateway.

Results: Within a week the team launched an official consortium gateway, allowing people across the country to access projections about mortality rates within different states and metropolitan areas. Used by CDC, White House, local policy makers. Over 400K visits from 180+ countries. Featured on CNN, New York Times. Additional external funding of $389,400 has been identified by the client.



College of Natural Sciences TEXAS

The University of Texas at Austin COVID-19 Modeling Consortium

PROJECTIONS PUBLICATIONS PEOPLE

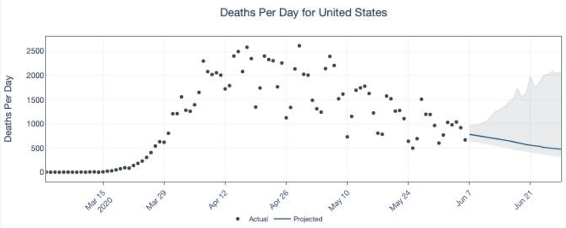
The University of Texas COVID-19 Modeling Consortium

COVID-19 Mortality Projections for US States

These graphs show both the reported and projected number of COVID-19 deaths per day across the US and for individual states. We use local data from mobile-phone GPS traces to quantify the changing impact of social-distancing measures on "flattening the curve."
Code syntax and daily updates of our forecasts are available on our UT-COVID GitHub repository.

Select your area in the dropdown menu below

Select State: United States **Select Metro Area:** Metropolitan Statistical Areas

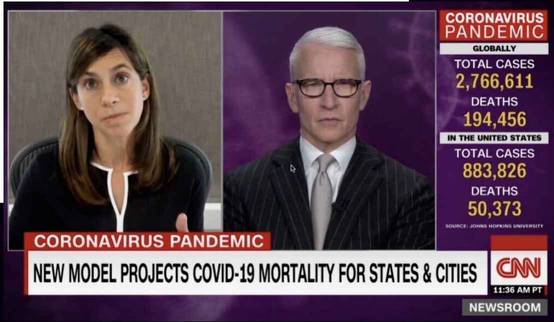


☐Show Log Scale ◉Daily ○Cumulative

* Click and drag on the plot to zoom-in and double-click to zoom-out completely. The icons at the top right of each graph will provide you with additional options.

Note: On May 7, 2020 we switched to using data on confirmed and probable COVID-19 deaths to make forecasts, rather than only confirmed deaths.

Key model assumptions: The model estimates the extent of social distancing using geolocation data from mobile phones and assumes that the extent of social distancing does not change during the period of forecasting. On June 2, 2020 we started reporting an ensemble forecast, combining predictions from our original model, based on "curve-fitting," and a new SEIR-based model.



GOOGLE COLAB

https://colab.research.google.com/

TACC

REU Community Coding Day 1

Let's do this!

PRESENTED BY:

Computational Thinking Objectives

The student will …

- Learn the about the concept of “computational thinking”
- Practice algorithm implementation through abstraction
- Learn about the concept of pseudo code

Back when mathematicians were computers and computers were calculators...

- Initially all programming was dedicated to translating math formulas.
- The work lead to the language FORmula TRANslation.

“Computational Thinking is the translation of ideas into computer code” ~Victor Eijkhout

Mathematical Thinking

- Number of people an elevator takes per day
- Speed (velocity) of an elevator
- Distribution of people in an elevator

Computational Thinking

- If there are X # of people expected to use elevators, how many should be installed?
- If someone at floor 0 presses the call button and there are available cars on floors 5 and 9, which car should respond?

The Process of Forming Logic (Think teaching a 3 year-old)

How would you tell a three(3) year old family member to get your keys out of the drawer in your room ?

Requirements, Logic, Algorithms, and Parameters

Requirements - what elements are needed before the job can be taken on

Logic - a system or set of principles underlying the arrangements of elements in a computer or electronic device so as to perform a specified task [an order in which to do a task]

Algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer [logic + calculations = algorithm]

Parameters - a limit or boundary that defines the scope of a particular process or activity [limits set on an algorithm = parameters]

Making a PB&J Sandwich

Requirements - what elements are needed before the job can be taken on

Parameters - a limit or boundary that defines the scope of a particular process or activity [limits set on an algorithm = parameters]

Logic - a system or set of principles underlying the arrangements of elements in a computer or electronic device so as to perform a specified task [an order in which to do a task]

Algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer [logic + calculations = algorithm]

Finding a definition in a dictionary

Requirements - what elements are needed before the job can be taken on

Parameters - a limit or boundary that defines the scope of a particular process or activity [limits set on an algorithm = parameters]

Logic - a system or set of principles underlying the arrangements of elements in a computer or electronic device so as to perform a specified task [an order in which to do a task]

Algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer [logic + calculations = algorithm]

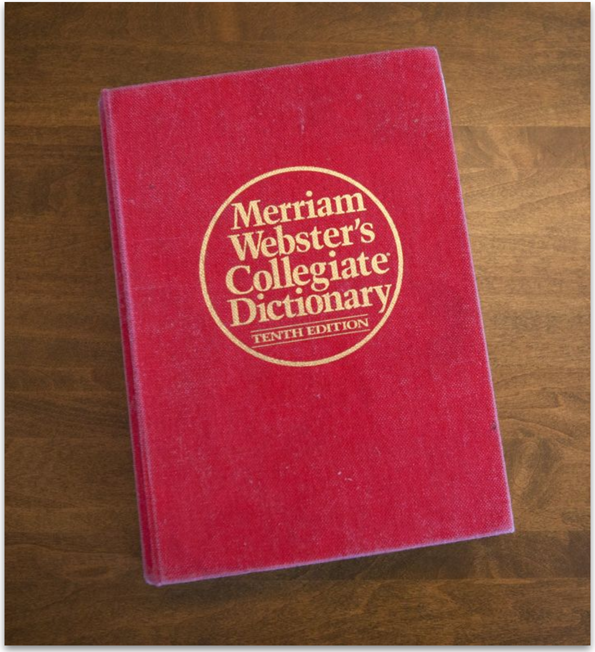
Given (Input):

- Dictionary
- Word (string)

Find (Output):

- Definition

Define your algorithm



Sorting

Requirements - what elements are needed before the job can be taken on

Parameters - a limit or boundary that defines the scope of a particular process or activity [limits set on an algorithm = parameters]

Logic - a system or set of principles underlying the arrangements of elements in a computer or electronic device so as to perform a specified task [an order in which to do a task]

Algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer [logic + calculations = algorithm]

Given:

A bag of potatoes

Problem:

Sort the bag of potatoes from smallest to largest

Algorithm:????

TACC

Think outside the "Box"

Requirements - what elements are needed before the job can be taken on

Parameters - a limit or boundary that defines the scope of a particular process or activity [limits set on an algorithm = parameters]

Logic - a system or set of principles underlying the arrangements of elements in a computer or electronic device so as to perform a specified task [an order in which to do a task]

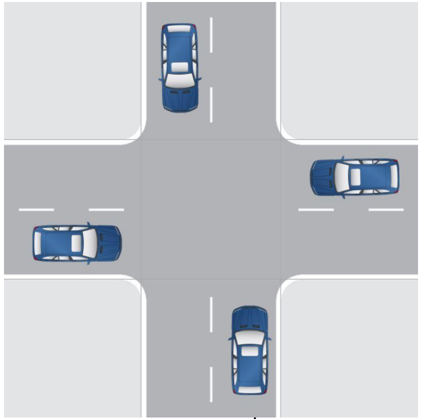
Algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer [logic + calculations = algorithm]

Problem:

4 automated cars come to an intersection at the same time.

Who goes first?

Algorithm: ???



What decisions did you make?

Requirements - what elements are needed before the job can be taken on

Parameters - a limit or boundary that defines the scope of a particular process or activity [limits set on an algorithm = parameters]

Logic - a system or set of principles underlying the arrangements of elements in a computer or electronic device so as to perform a specified task [an order in which to do a task]

Algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer [logic + calculations = algorithm]

What was the last meal you ate?

What were the defining parameters on why you chose that meal?

Where do we go?
Where do we go now?

Look at each problem you are going to tackle, and figure out the requirements - what is needed to solve? Figure out the logic on how to solve it, and apply the algorithm.

TACC

Think about a Scientific Process

Let's meet Joe.

Joe might get sick.

Joe will be sick for 5 days.

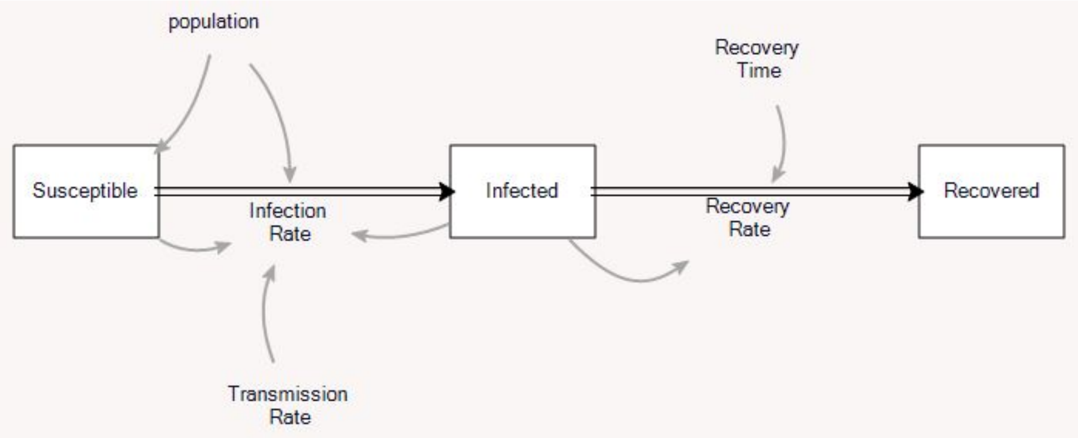
After 5 days, Joe gets better.

Once Joe gets better, Joe can no longer get sick.

How would we "code" Joe?

"Hi I'm Joe"

The SIR Model



Task 1 - Code Joe

Variables to hold data

Mathematical Operations to do math :)

Conditionals to make decisions

Loops to repeat our process

Functions/Subroutines to reuse code

Objects or Classes to define our "things"

Let's meet Joe.

Joe might get sick.

Joe will be sick for 5 days.

After 5 days, Joe gets better.

Once Joe gets better, Joe can no longer get sick.

Let's "code" Joe.

Task 2 Code Joe and Jane

We met Joe.

Joe has a friend, Jane

If Joe gets sick, Jane might get sick.

Modify your code, so when Joe gets sick that triggers Jane to roll a random number to see if Jane gets sick.

Loop through your code until both Joe and Jane get sick and they each get better.

"Hi I'm Jane and I unfortunately know Joe"

TACC

Object Oriented Python

What's an Object?

Object-oriented programming is a programming paradigm that provides a means of structuring programs so that properties and behaviors are bundled into individual objects.

Hackathon Attendees and Je'amime (and me):   "HUH?!"

Object Oriented Python

What's an Object? - Let's try this again

Object-oriented programming allows you to code up "things" that have ***properties/attributes*** (variables) and ***behaviors/methods*** (functions)

These "things" allow us to create reusable code and allows us mimic real "things"

Object Oriented Python

Hackathon Attendees and Je'aime (and me):   "AHHH!"

Object Oriented Python

How do we go about doing this?

Classes

A class is a blueprint for how something should be defined.

Object Oriented Python

How do we go about doing this?

Classes

A class is a blueprint for how something should be defined.

Let’s design our ***Person* Class**

Object Oriented Python

We need another concept.

the *instance?*

The class is the blueprint, an ***instance*** is an object that is built from a class and contains real data.

The **class** would be a form, the **instance** is the form filled out.

Object Oriented Python

We need a couple bits of syntax.

the *self*

The ***self*** parameter is a reference to the current *instance* of the class, and is used to access variables that belongs to the class.

Object Oriented Python

We need a couple bits of syntax.

the __*init*__

When you create a new object of a class, Python automatically calls the __init__() method to initialize the object's attributes.

Task 2 Code Joe and Jane

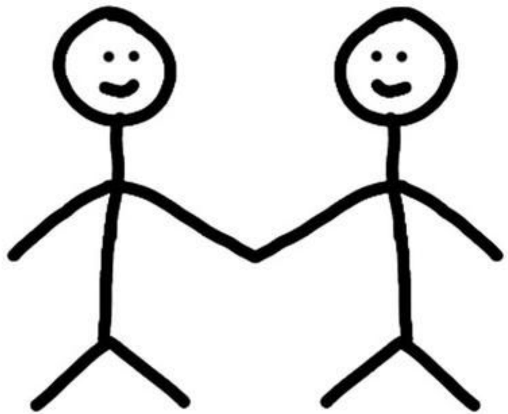
We met Joe.

Joe has a friend, Jane

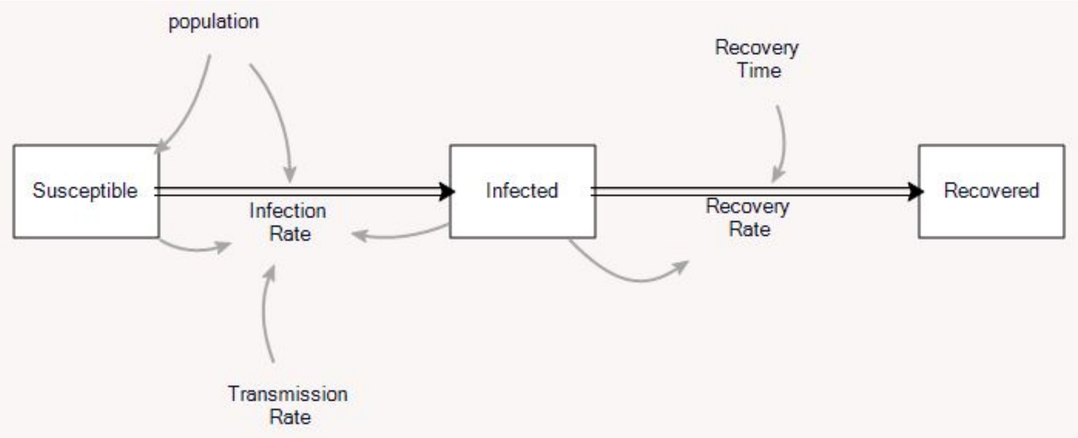
If Joe gets sick, Jane might get sick.

Modify your code, so when Joe gets sick that triggers Jane to roll a random number to see if Jane gets sick.

Loop through your code until both Joe and Jane get sick and they each get better.



The SIR Model



Task 1 - Code Joe

Variables to hold data

Mathematical Operations to do math :)

Conditionals to make decisions

Loops to repeat our process

Functions/Subroutines to reuse code

Objects or Classes to define our "things"

Let's meet Joe.

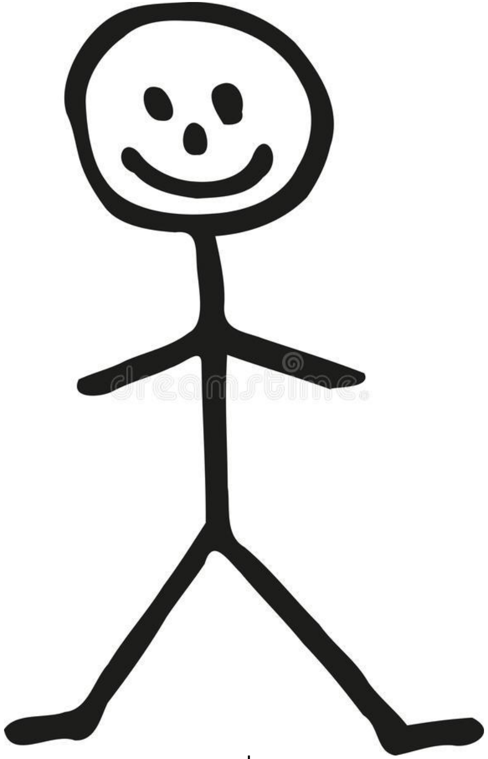
Joe might get sick.

Joe will be sick for 5 days.

After 5 days, Joe gets better.

Once Joe gets better, Joe can no longer get sick.

Let's "code" Joe.



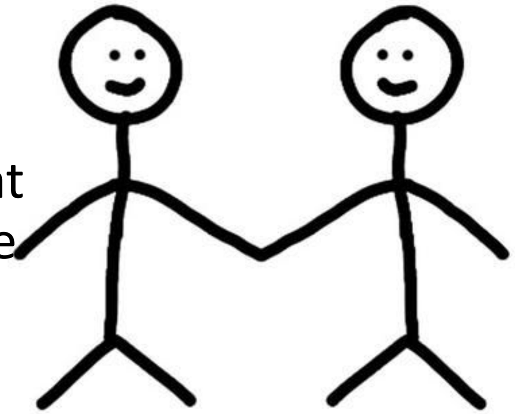
Task 2 Code Joe and Jane

We met Joe.

Joe has a friend, Jane

If Joe gets sick, Jane might get sick.

Modify your code, so when Joe gets sick that triggers Jane to roll a random number to see if Jane gets sick.



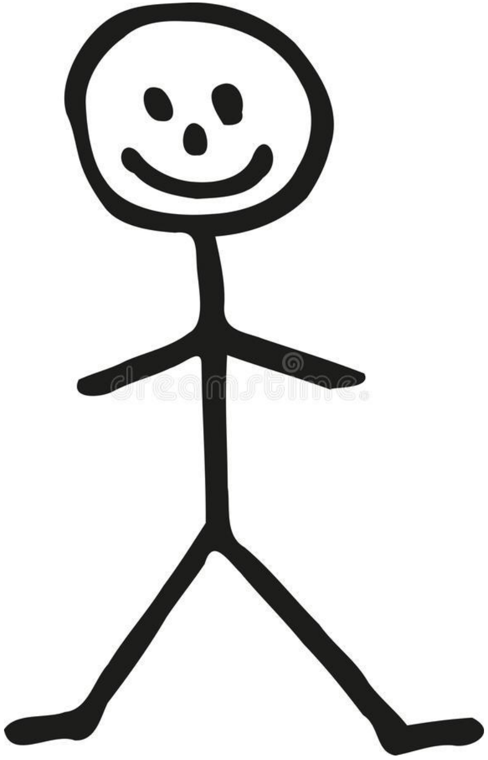
Loop through your code until both Joe and Jane get sick and they each get better.

Task 3 Code A Person Object.

Joe, is now of type *Person*. In our "world":

Joe can get infected
Joe can get sick for certain number of days
Joe can get better - once better, let's assume Joe can't ever get sick again
Joe can get vaccinated - let's assume that the vaccine is awesome, so once again, once vaccinated, Joe can't get sick again.

Code a Person object with the above criteria, and "instantiate" Joe. Now, go back to Task 1, and make Joe the Person Object, sick for 5 days using the same algorithm.



Task 4 Code the Population

We now have a Person object, we now need a Population object

- The Population is basically a Vector of Person objects
- The Population Object will need methods to add a person, return the number of Persons, return the number of sick Persons, well Persons, inoculated Persons
- Your Population will start out at 1000 Persons

Task 5 Code One Day of Interactions

The Person also interacts with a number of other Persons each day

The Person Object will have a "Interactions" Vector
For each Person, randomly fill the Interactions Vector with 10 other Persons.
The Interactions do not need to be reciprocal, meeting if Person P2 shows up in Person P1's Interaction Vector, P1 doesn't necessarily have to show up in P2's Interaction Vector (it should! but we should keep our simulation simple)

Task 6 Code the Population
with full Interactions and a propagating Contagion

We now have a Person object and a Population object. Remember, once a person gets better, they can no longer get sick, they are inoculated

- Set a population of 1000 people
- Introduce Patient Zero
- This is your first sick person
    - Day 1. Patient Zero interacts with others, those other may now get sick as well.
- As each Day passes
    - sick people are interacting with well people, causing the virus to spread
    - each day print out the number of sick people, and the number of inoculated people
- The simulation ends, when there are no more sick people.

Task 7 Code an Intervention

You see that the Peak in a 40,000 people community is very high

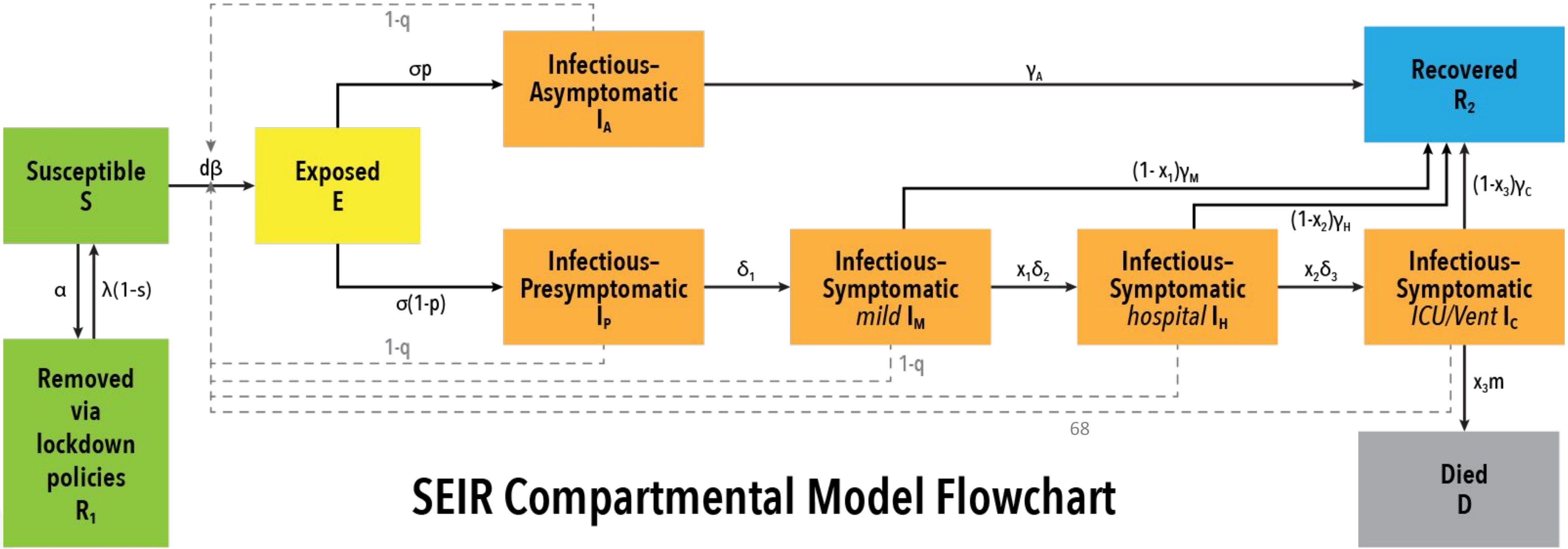
What are the benefits of Masking?

How do Masks affect the infection rate?

Your challenge: Modify your models so that a certain percentage of your Population wears masks, and model the Mask benefits.

TACC

The Covid SEIR Model



SEIR Compartmental Model Flowchart