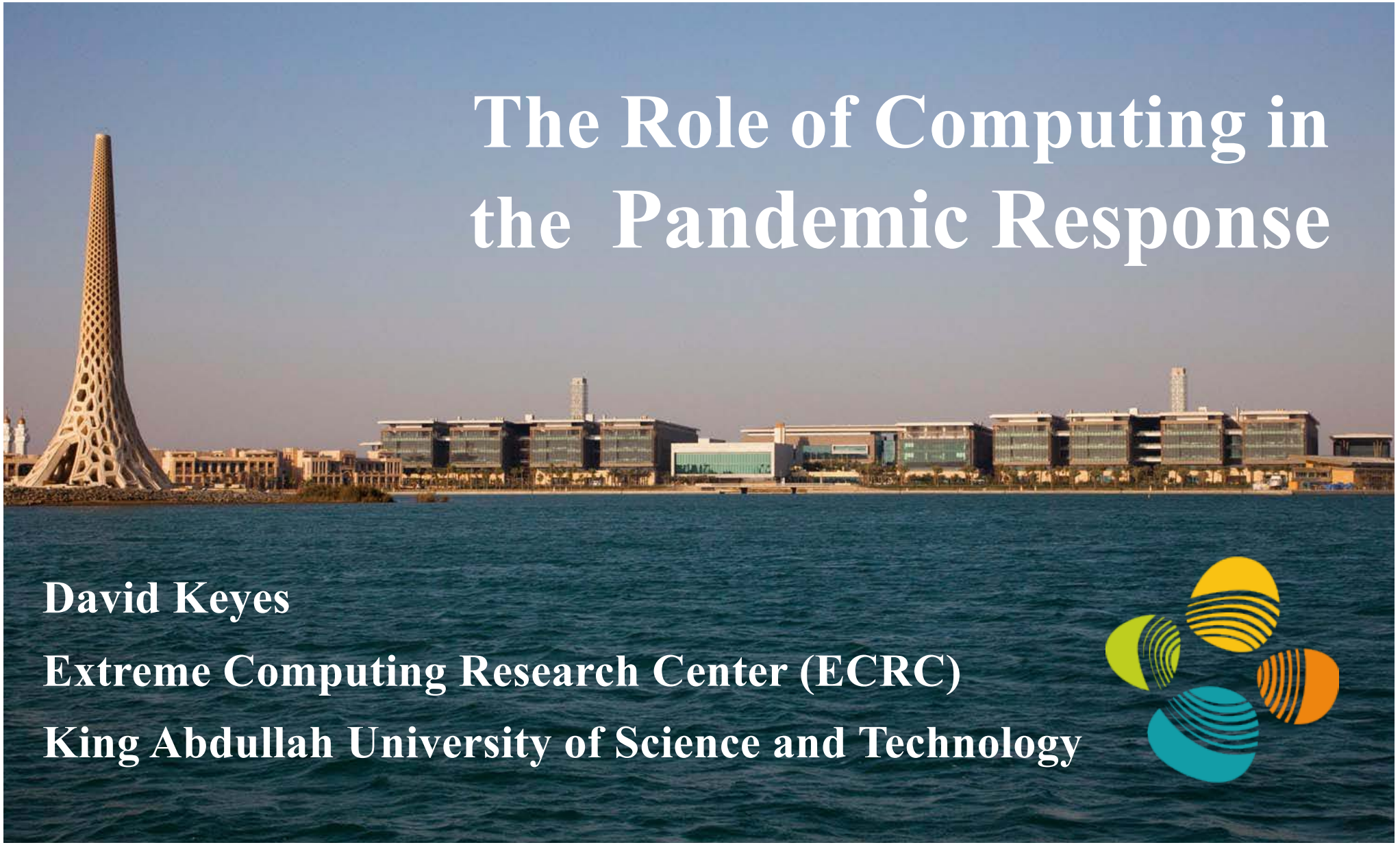
siam 2021 | Conference on Computational Science and Engineering

MS 12 & 78



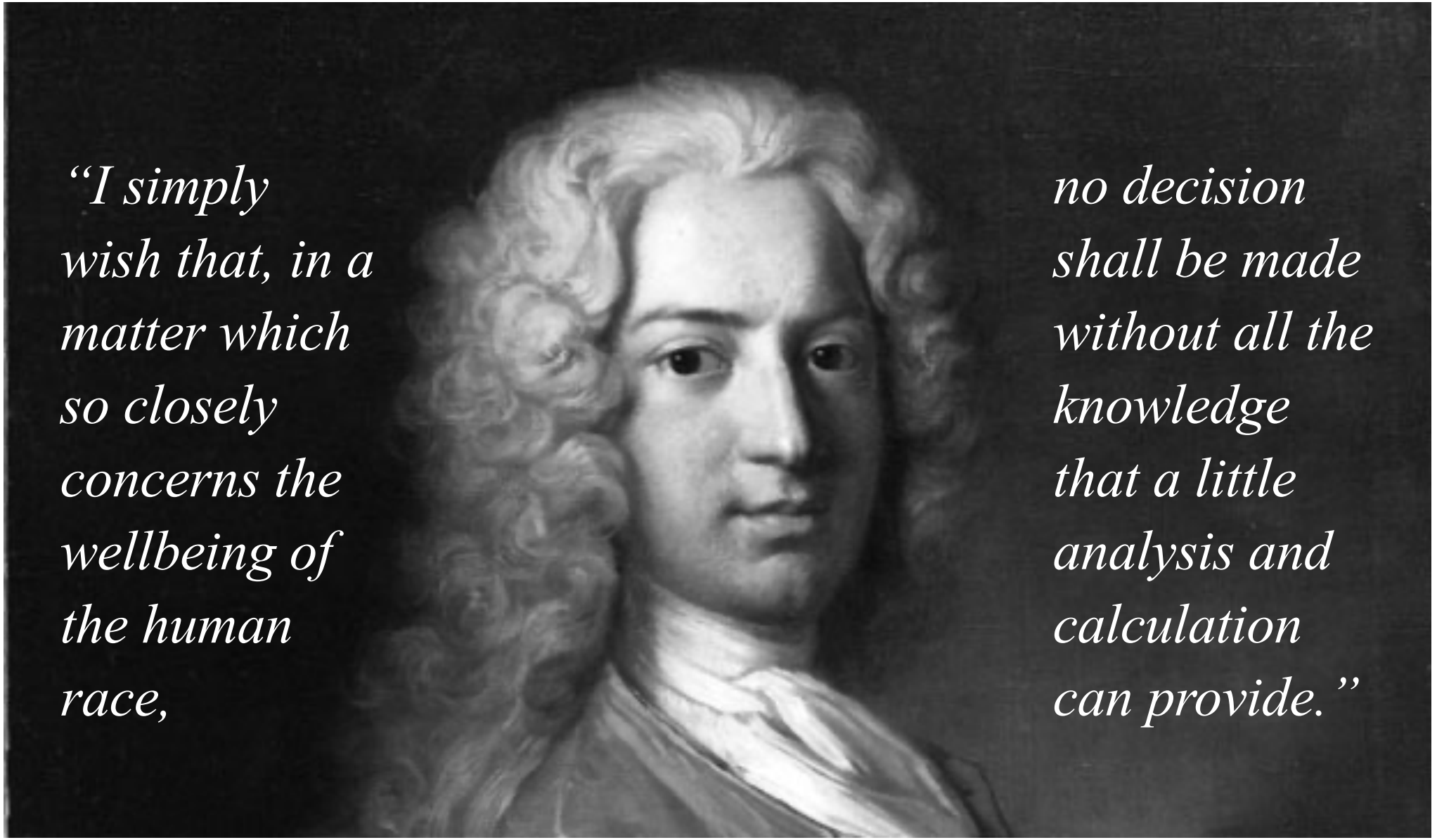
The Role of Computing in the Pandemic Response

David Keyes

Extreme Computing Research Center (ECRC)

King Abdullah University of Science and Technology

… Paris 1760 …


"I simply wish that, in a matter which so closely concerns the wellbeing of the human race,
no decision shall be made without all the knowledge that a little analysis and calculation can provide."

Daniel Bernoulli's analysis of smallpox

Essai d'une nouvelle analyse de la mortalité causée par la petite vérole & des avantages de l'inoculation pour la prévenir

(An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it)

$$-ds = \frac{sdx}{n} - s\frac{d\xi}{\xi} - \frac{ssdx}{mn\xi}$$

x: time
s: susceptibles
ξ: survivors
$1/n$, $1/m$: rate constants

presented at
The Royal Academy of Sciences in Paris
on 16 April 1760

published in
Histoire et Mémoires de l'Académie Royale des Sciences de Paris,
1766, pp. 1-45

… March 2020 …

- COVID19 HPC Consortium established by DOE, NSF, industry, and academia in the US
- COVID19 Fast Track established by PRACE in the EU

… April 2020 …

- Fugaku brought into service a year early by RIKEN and MEXT in Japan expressly to fight COVID19

… June 2020 …

ISC 2020 DIGITAL AGENDA, JUNE 22 – 25

ISC SHAPING TOMORROW High Performance

from 5 pm Focus Session **Covid 19 & HPC/AI**

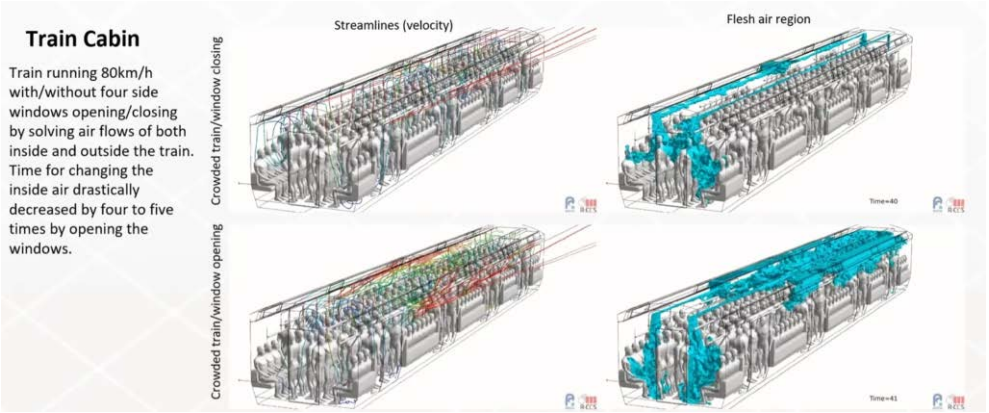
Presentations:

Japan and Fugaku's Fight Against the COVID-19 in HPC ★
Speaker: Satoshi Matsuoka

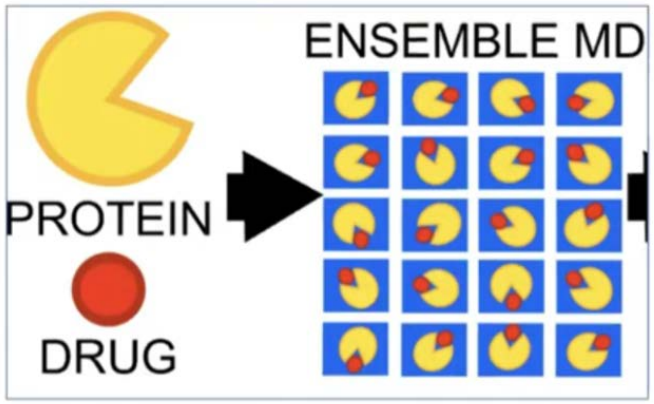
Supercomputing, COVID-19 and the Transformation of Medicine ★
Speaker: Peter Coveney

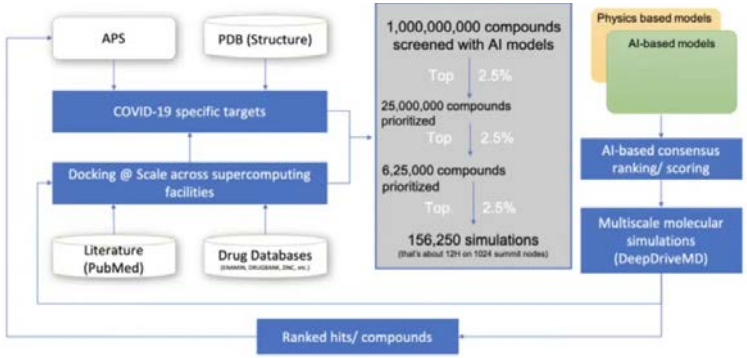
Overview of HPC and AI Computing for COVID-19 in the US ★
Speaker: Rick Stevens

… June 2020 …



HPC wire





… June 2020 …

"COVID19 has had considerable positive impact on CSE research but not *vice versa*."

"COVID19 understanding and mitigation is still essentially a wet lab and clinical story."

"Where has CSE had an impact beyond research frontiers – in policy support, diagnosis, or therapy?"

… September 2020 … *

"Recent results highlight the importance of **computational science as a guiding tool that helps in uncovering the structure and mechanics of the viral spread as well as drastically reduces the number of candidate treatments that need laborious laboratory testing**. From transformative medical advances to drug discovery driven by massively parallel supercomputing, this minisymposium will span a range of practical approaches to the ongoing crisis."

***from the abstract of MS 12 & 78**

… March 2021 …

Report Card
Subject
Grade
Notes
Visualization
A
virus geometry
Molecular dynamics
A
virus mechanisms
Fluid-dynamical transport
A
safety protocols
Diagnosis
A
clinical accuracy
Drug repurposing
I
efficient trials
Tracking spread & variants
I
understanding scope
Forecasting spread
I
planning resources
Optimizing response
I
setting policy

Summary (1): CSE has been good for COVID19

- Allowed scientists to visualize hundred million-atom biomolecular clusters in far greater detail than any laboratory imaging system
- Allowed scientists to understand the ACE2 “attack” configuration of the spike protein with its glycan “kickstand” supports
- Allowed health & safety experts to understand convective and diffusive transport of the virus in rooms and passenger vehicles
- Gave clinicians new AI-assisted CT scan interpretation
- Identified hundreds of potential new drugs computationally and launched dozens into experimental pipelines and clinical trials
- Created computational infrastructure and workflows relevant to future pandemics
- Enabled public policy makers to smartly dial the balance between losing lives and suspending livelihoods

Summary (2): COVID19 has been good for CSE

- Pushed #1-ranked system into broad use a year ahead of schedule
- Created new intercontinental HPC consortia
- Led to sharing of data on unprecedented scale
- Created new computer-to-laboratory pipelines
- Focused new attention on the role of preprint archives
- Accelerated injection of AI into scientific and clinical workflows
- Accelerated development of new algorithms in MD
- Accelerated development of new models in epidemiology, population-stratified beyond the classical S-E-I-R, and Bayesian methods in their parameter fitting
- Brought billions of readers and viewers to the precipice of computational modeling and machine learning, “looking in” from popular media accounts; has inspired future scientists

An example of each of the eight campaign types*

Contribution	Contributor	Release date	AI intensity	Comp intensity	Data intensity	Repro-ducibility
All-atom Visualization	Ivan Viola, KAUST	Apr'20	none	med	large	code
CT Scan Diagnosis	Xin Gao, KAUST	May'20	high	med	med	code & data
Forecasting Spread	Paula Moraga, KAUST	Jun'20	none	low	low	code & data
Optimizing Response	David Ketcheson, KAUST	May'20	none	low	low	code
Tracking Variants	Intikhab Alam, KAUST	Feb'21	none	low	large	web rsrc
Fluid-dyn. Trans.	Ville Vourinen, Aalto	May'20	none	med	low	code
Molecular Dyn.	Rommie Amaro, UCSD	Nov'20	high	high	large	code & data
Drug Disc./Repurp.	Hari Arthanari, Harvard	Feb'21	none	high	large	code, data, web rsrc

* Many of the best examples globally are in the other presentations in MS 12 & 78. This line-up is chosen to complement, featuring some local sources.

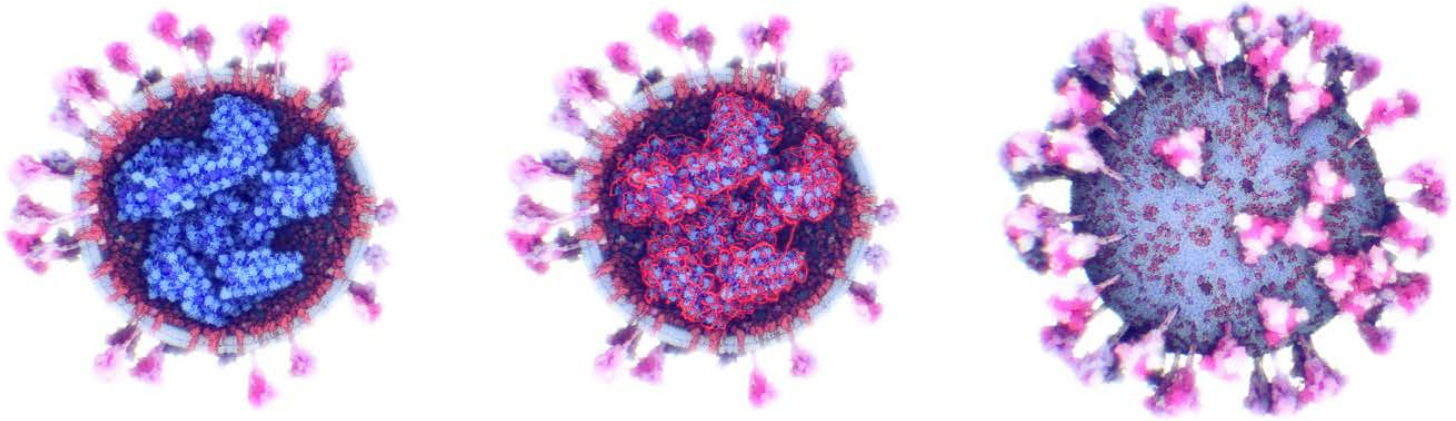
Where to find these slides

https://cemse.kaust.edu.sa/sites/default/files/2021-03/ms78keyes.pdf

All-atom Visualization (I. Viola, KAUST)

Modeling in the Time of COVID-19:
Statistical and Rule-based Mesoscale Models

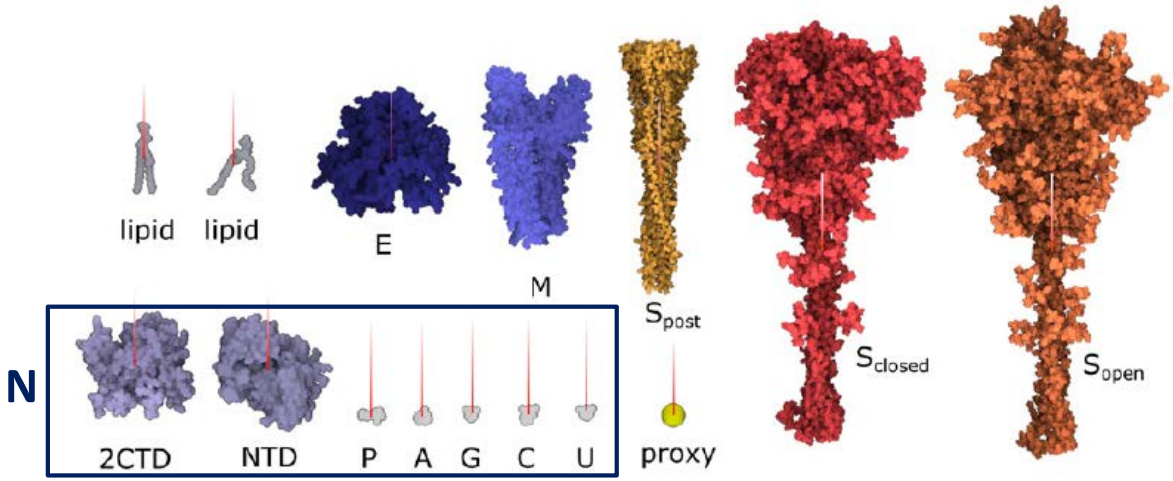
Ngan Nguyen, Ondřej Strnad, Tobias Klein, Deng Luo, Ruwayda Alharbi,
Peter Wonka, Martina Maritan, Peter Mindek, Ludovic Autin, David S. Goodsell, Ivan Viola

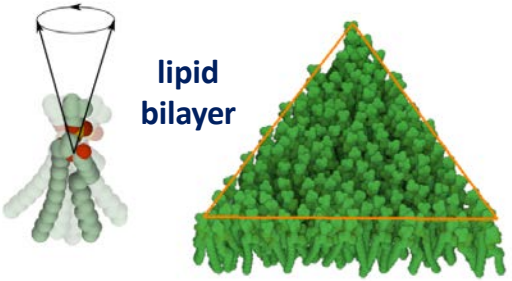


- **Pipeline for visualization of mesoscale biological structures**
  - **microscopic images → segmentation → statistical modeling → user-defined rules → 3D model, iteratively refined by modifying the rules**
- **Premiums placed on:**
  - **scientific relevance, rapidity, reusability, revisionability, controllable precision**
- **Applied in April to SARS-CoV2 to provide early, progressively updated viz**

All-atom Visualization (I. Viola, KAUST)

- **Exploit rules-based structure (23 rules)**
- **Basis of structural elements (15, shown here)**





S (spike): receptor binding domain (RBD) protein attacks host cell

M (membrane): packages viral genome as virus buds from infected cell's surface

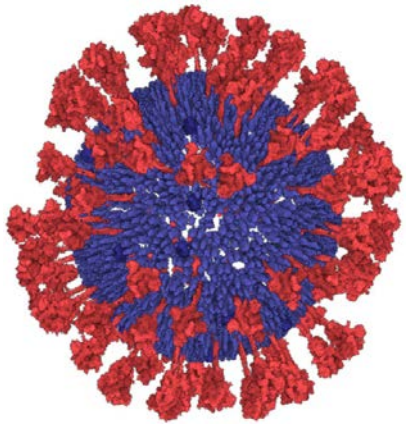
E (envelope): forms ion pore through the membrane

N (nucleocapsid): scaffold for the single-stranded RNA inside virion

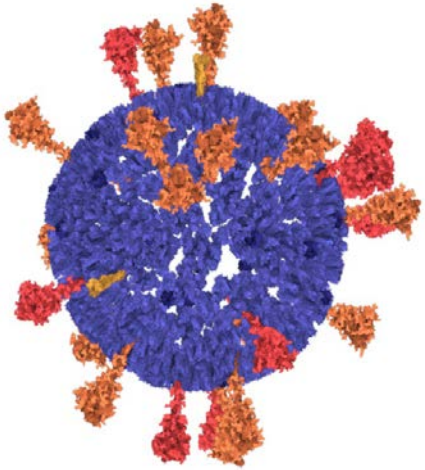
- **Assemble mesoscale objects through learned rules guiding number, placement, orientation, and interaction**
- **Augment structure definition through 3D interactive modeling**
- **Such procedural modeling has long tradition in computer graphics (vegetation, clouds, buildings, etc.)**
  - **combined here with real-time visual programming for user productivity**

All-atom Visualization (I. Viola, KAUST)

Revisionability: regularly updated with new research

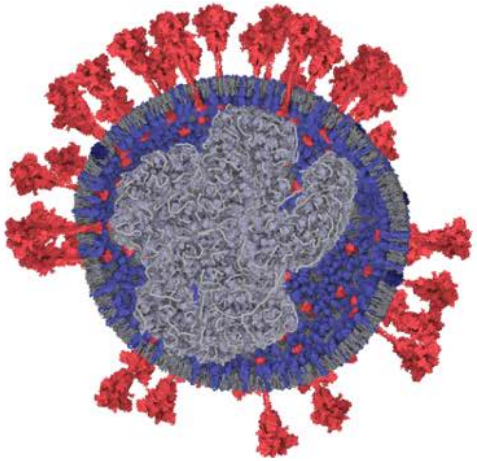


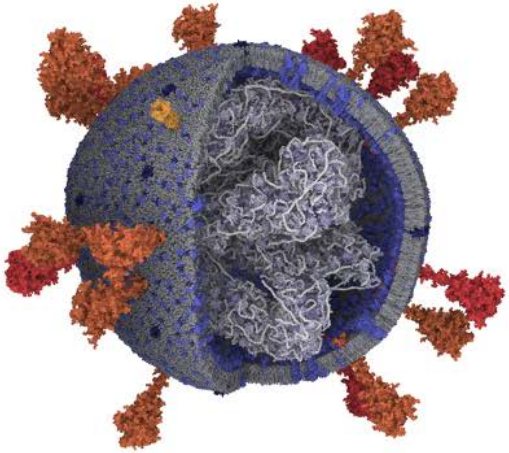
April 2020

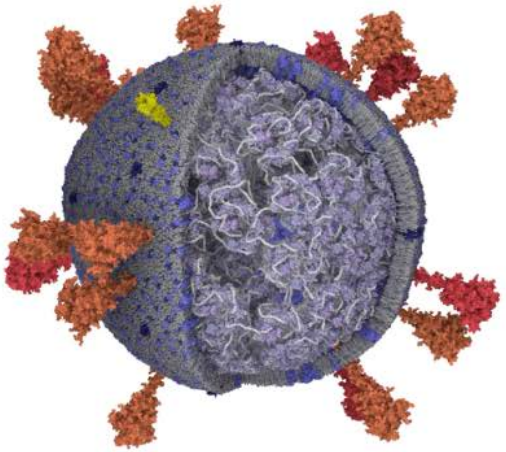


June 2020

September 2020







All-atom Visualization (I. Viola, KAUST)

New: modeling based directly on imaging [video, 1:31]

- **Begins with a 3D image of a field of viruses rendered by "tilt series" cryo-ET microscopy**
- **Continues with the "escape" from a plane of the image of a procedurally rendered COVID19 virus**
- **Shows the assembly of the virus model from protein components**
- **Finally, shows a 3D field of viruses with shapes and spacing inherited directly from the micrograph**

All-atom Visualization (I. Viola, KAUST)

Besides making the invisible visible to virologists, sciviz can refine fanciful COVID19 images in the popular press

Mail Online is the most visited English-language newspaper website with about 50% more readers than The New York Times

All-atom Visualization (I. Viola, KAUST)

The KAUST-Nanographics® visualization of a virus attacking the epithelium ran on the cover of the fall semester 2020 issue of *KAUST Impact.* Research translation begins at home.

CT Scan Diagnosis (X. Gao, KAUST)

2638 IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 39, NO. 8, AUGUST 2020

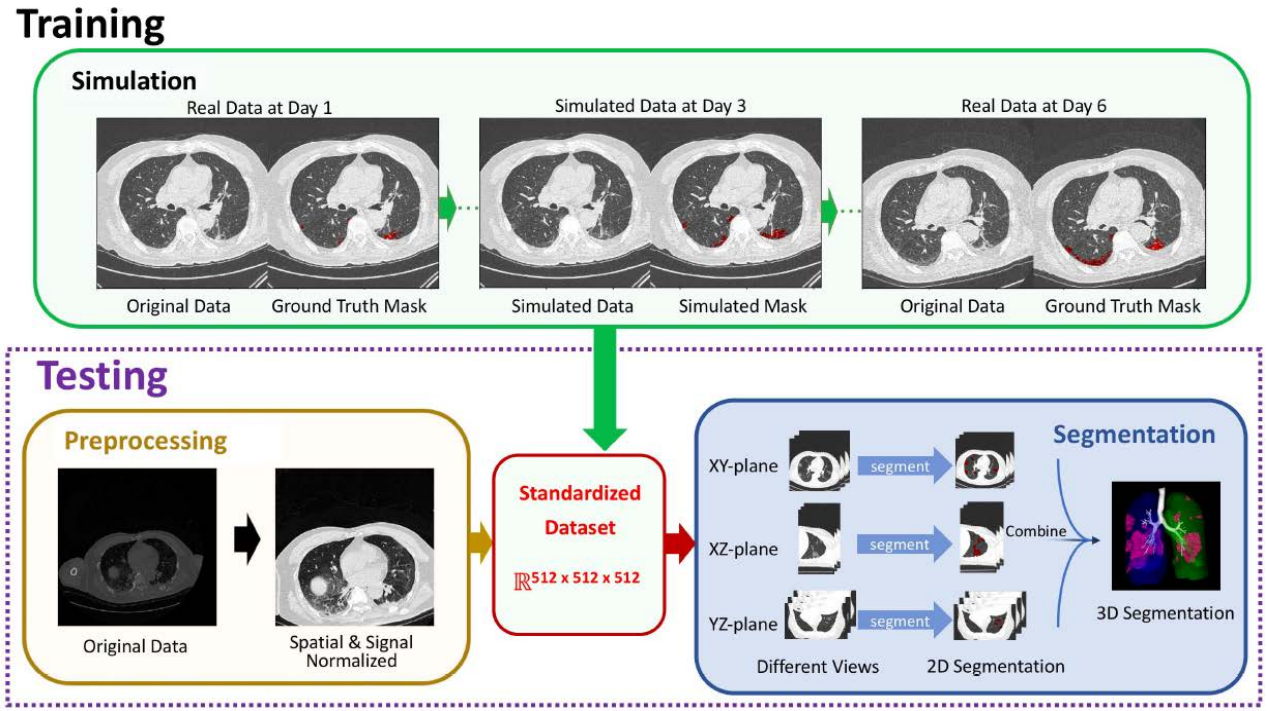
A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis

Longxi Zhou, Zhongxiao Li, Juexiao Zhou, Haoyang Li, Yupeng Chen, Yuxin Huang, Dexuan Xie, Lintao Zhao, Ming Fan, Shahrukh Hashmi, Faisal Abdelkareem, Riham Eiada, Xigang Xiao, Lihua Li, *Member, IEEE*, Zhaowen Qiu, *Member, IEEE*, and Xin Gao, *Member, IEEE*

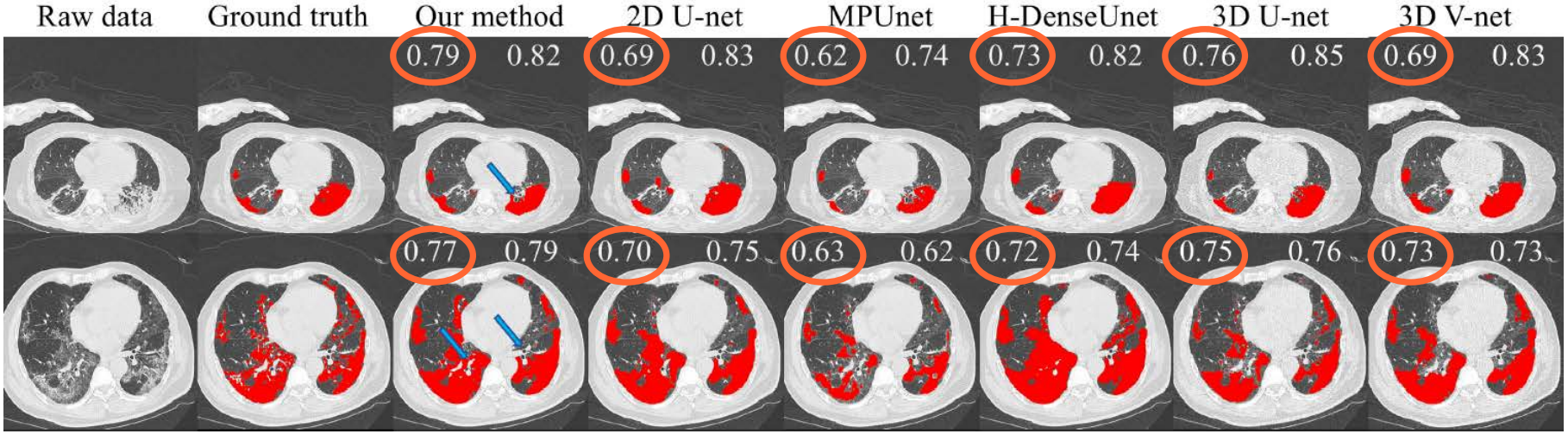
- **Nucleic acid-based testing is the gold standard for an RNA-carried disease but has high false negative rate, particularly in early stages**
- **Automatic, rapid, accurate, and machine-agnostic method demonstrated to segment and quantify infected regions in CT scans, with two innovations:**
  - **alleviate training data scarcity issue**
    - ***fit shots of same-patient data measured at different time points and augmented with simulated data in between scan times***
  - **solve large-scene, small-object problem**
    - ***decompose 3D segmentation problem into three 2D problems, reducing model complexity by an order of magnitude and improving segmentation accuracy***

CT Scan Diagnosis (X. Gao, KAUST)

Pipelines:



Results:



CT Scan Diagnosis (X. Gao, KAUST)

- **AI-interpreted X-ray and CT scans not clinically new**
  - **superior to radiologists in many cases, already**
- **COVID-19 CT scans achieve 97% sensitivity in those confirmed with positive RT-PCR testing, but detect about 30% more positive cases than RT-PCR**
- **Based on Convolutional Neural Networks, which are strong in classifying 2D images**
- **Extracting multiple 2D images often superior to 3D (called "2.5D")**
  - **current program stacks 2D images in each of three orthogonal view angles**
  - **fewer hyperparameters, smaller memory footprint – faster to train**
  - **real-time inference vs. much longer times for 3D inference**
- **Trained over multiple CT machines with data from multiple hospitals**
- **Single-patient data augmented by time-elapsed simulated data**
- **Pre-processing maps into machine-agnostic standard space, 512^3, representing volume of $334 \times 334 \times 512$ mm^3 per voxel**
- **Especially effective in early onset**

CT Scan Diagnosis (X. Gao, KAUST)

STAGE-SPECIFIC *Segmentation* PERFORMANCE

Methods	The *Harbin* dataset		
	Early	Progressive	Severe
Percentage	73.1%	18.7%	8.2%
Ours	**0.764**	**0.823**	**0.875**
2D U-net	0.508	0.703	0.755
H-DUnet	0.526	0.788	0.803
MPUnet	0.370	0.618	0.766
3D U-net	0.551	0.802	0.820
3D V-net	0.581	0.796	0.824

The best performer under each criterion is in bold. The performance is shown in the form of average dice.

"The model was fast to use as each case took less than a minute to process. This will make an important contribution to chest imaging, especially during the current pandemic."

Dr. Riham Eiada, King Faisal Specialist Hospital

Forecasting Spread (P. Moraga, KAUST)

Wellcome Open Research

Wellcome Open Research 2020, 5:117 Last updated: 05 JAN 2021

Check for updates

RESEARCH ARTICLE

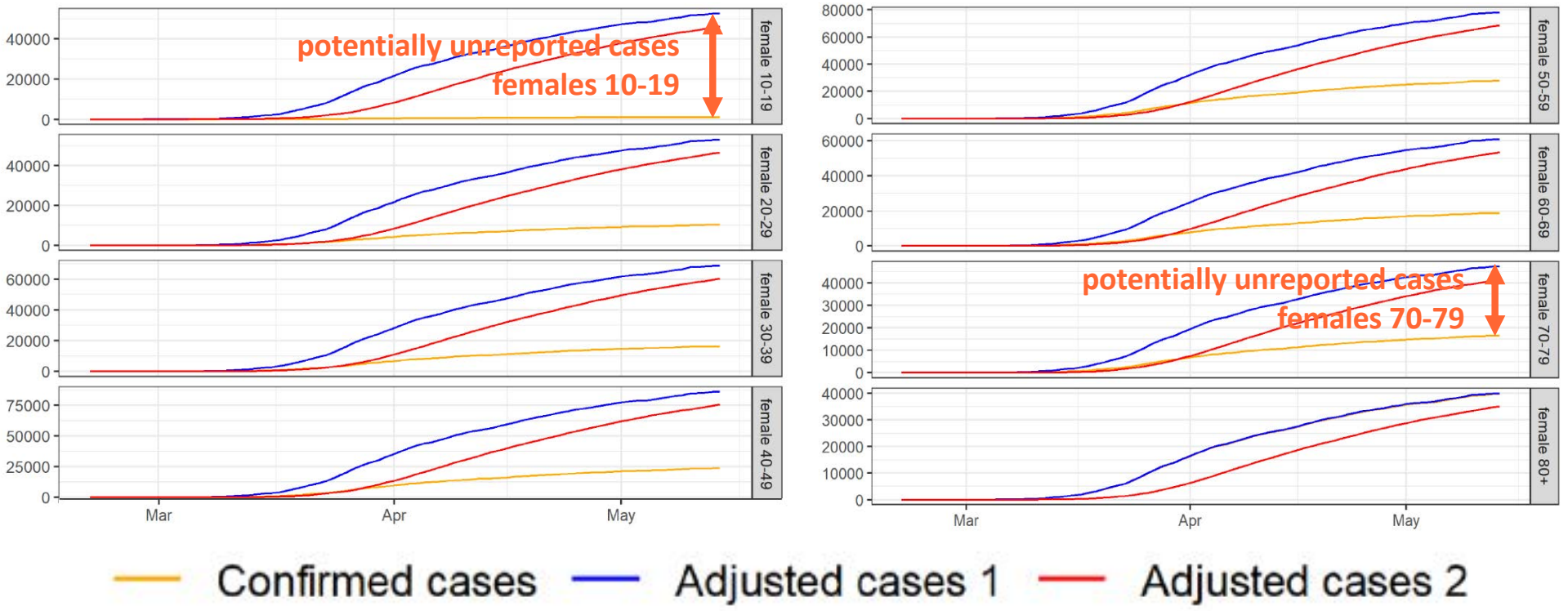
Assessing the age- and gender-dependence of the severity and case fatality rates of COVID-19 disease in Spain [version 1; peer review: 1 approved]

Paula Moraga[1], David I. Ketcheson[2], Hernando C. Ombao[2], Carlos M. Duarte[3]

[1]Department of Mathematical Sciences, University of Bath, Bath, Somerset, BA2 7AY, UK
[2]Computer, Electrical, and Mathematical Sciences & Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[3]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

- **Public health planners must infer actual infected numbers and predict future hospitalizations and fatalities from test and fatality data that preferentially captures those who present symptoms or are in vulnerable groups.**
- **Populations must be partitioned by age, gender, and other significant factors and rates adjusted based on assumptions that must be uncertainty quantified.**
- **Rates must also be adjusted to account for bias due to varying delay between disease onset and death.**

Forecasting Spread (P. Moraga, KAUST)



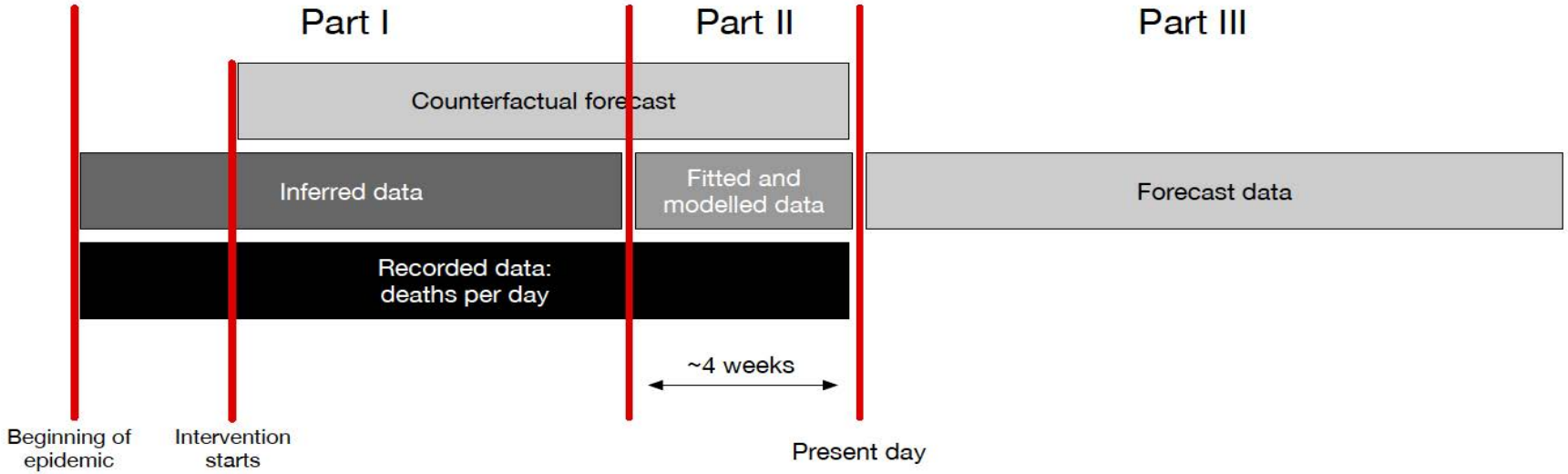
"This addresses all of the difficulties in calculating Case Fatality Rates during an outbreak and provides detailed estimates for Spain."

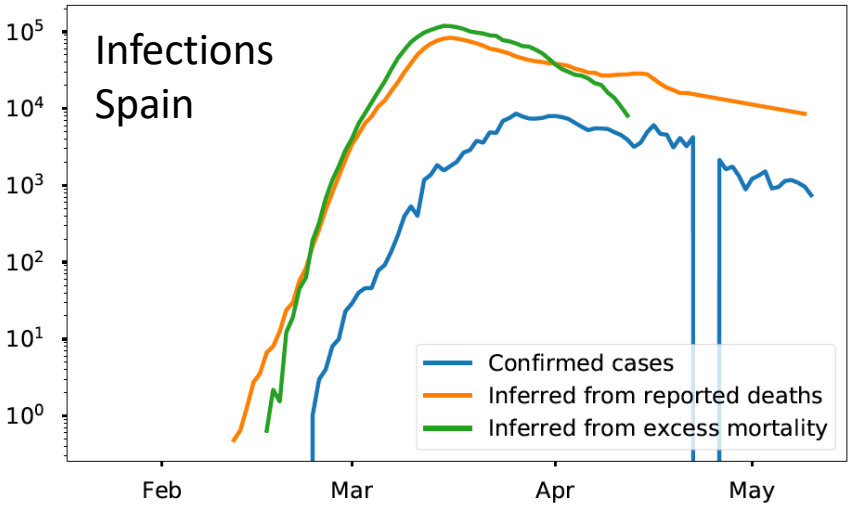
Timothy W. Russell, Dept. of Infectious Diseases
London School of Hygiene and Tropical Medicine

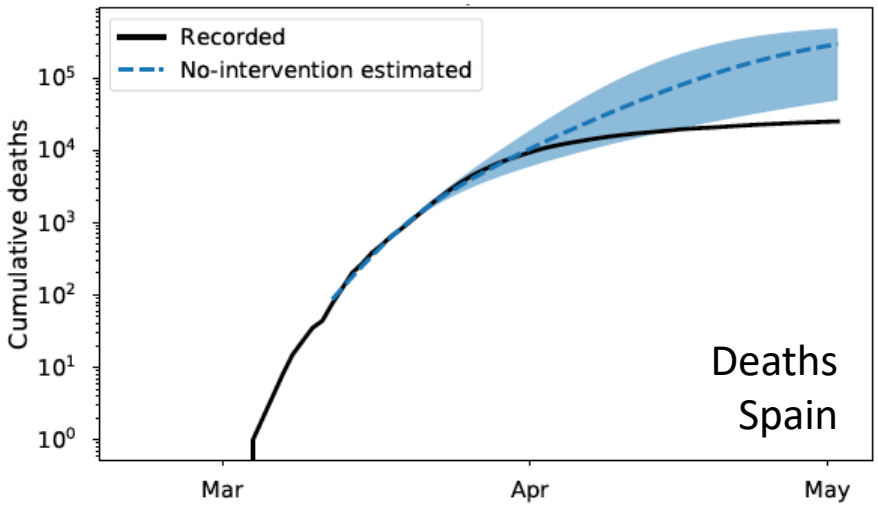
Forecasting Spread (P. Moraga, KAUST)

Because testing is highly selective, statistically corrected mortality data is employed.

Forecasts are run forward, and estimates for no intervention are run backward.

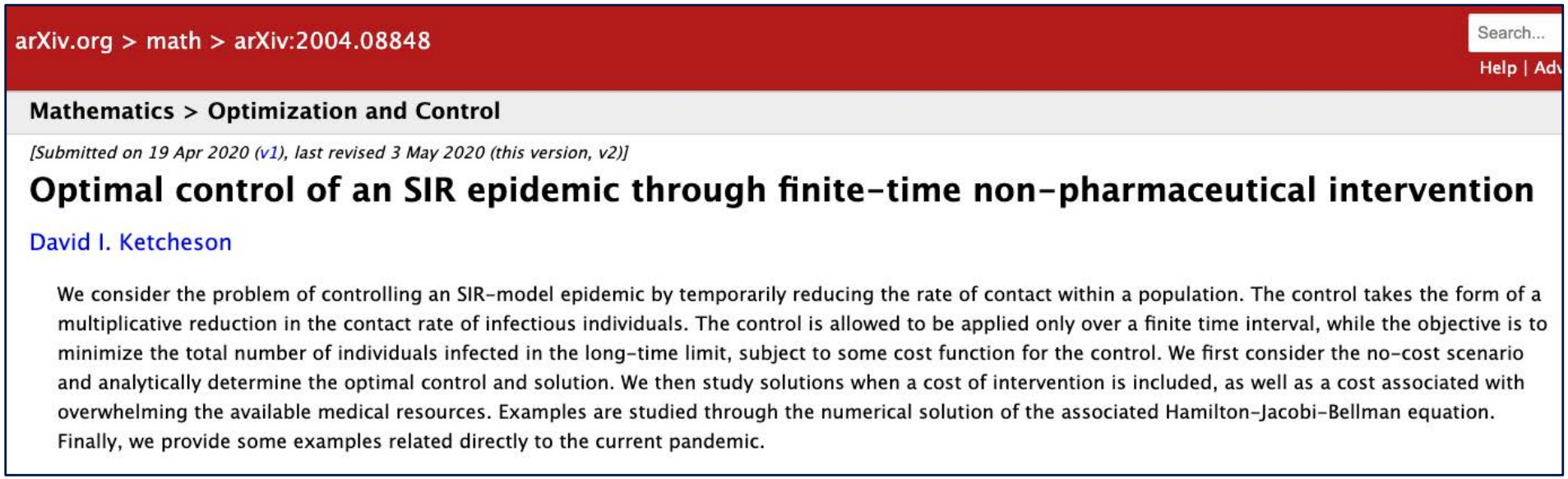






NB: vertical scales logarithmic

Optimizing Response (D. Ketcheson, KAUST)



arXiv.org > math > arXiv:2004.08848

Search...

Help | Adv

Mathematics > Optimization and Control

[Submitted on 19 Apr 2020 (v1), last revised 3 May 2020 (this version, v2)]

Optimal control of an SIR epidemic through finite-time non-pharmaceutical intervention

David I. Ketcheson

We consider the problem of controlling an SIR-model epidemic by temporarily reducing the rate of contact within a population. The control takes the form of a multiplicative reduction in the contact rate of infectious individuals. The control is allowed to be applied only over a finite time interval, while the objective is to minimize the total number of individuals infected in the long-time limit, subject to some cost function for the control. We first consider the no-cost scenario and analytically determine the optimal control and solution. We then study solutions when a cost of intervention is included, as well as a cost associated with overwhelming the available medical resources. Examples are studied through the numerical solution of the associated Hamilton-Jacobi-Bellman equation. Finally, we provide some examples related directly to the current pandemic.

- **Given an NPI (e.g., lockdowns, masking, etc.) of finite duration, when it is best applied to minimize the asymptotic number of deaths?**
- **With and without cost (as an additive term in the cost function)?**
- **With and without constraints (e.g., hospital capacity)?**
- **Formulated as a Hamilton-Jacobi-Bellman problem**
- **Results:**
  - **exact specification of the time that intervention should begin in the case of no cost**
  - **numerical explorations with nonzero cost and constraints, using COVID19 params**

Digression: S-I-R model (1927)

Proc. Royal Society London, Series A **115**(772), pp. 700-721

A Contribution to the Mathematical Theory of Epidemics.

By W. O. Kermack and A. G. McKendrick.

(Communicated by Sir Gilbert Walker, F.R.S.—Received May 13, 1927.)

(From the Laboratory of the Royal College of Physicians, Edinburgh.)

from the consideration of the special case in which ϕ and ψ are constants κ and l respectively.

In this case the equations are

Susceptibles: $\frac{dx}{dt} = -\kappa xy$

Infecteds: $\frac{dy}{dt} = \kappa xy - ly$

Removed (recovered/deceased): $\frac{dz}{dt} = ly$

(29)

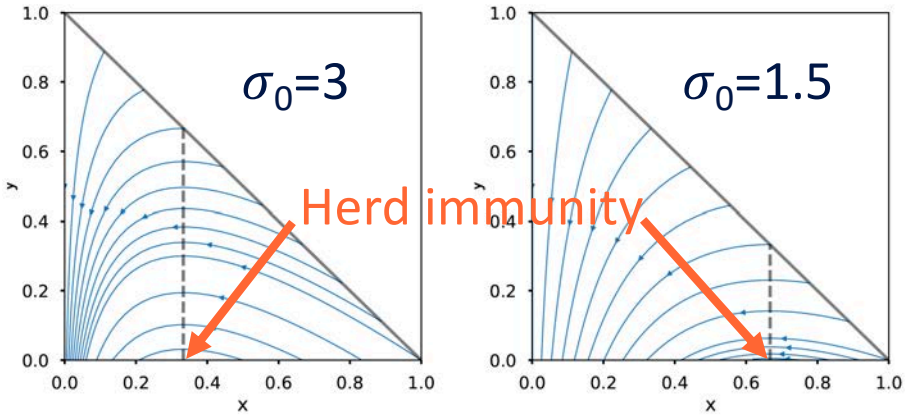
and as before $x + y + z = \mathrm{N}$.

Optimizing Response (D. Ketcheson, KAUST)

Non-controlled ODEs

$$\text{S: } x'(t) = -\gamma\sigma_0 y(t)x(t)$$

$$\text{I: } y'(t) = \gamma\sigma_0 y(t)x(t) - \gamma y(t)$$



Control problem:

Given $(x_0, y_0) \in \mathcal{D}, \sigma_0 > 0, T > 0,$

choose an admissible control $\sigma(t) : [0, T] \to [0, \sigma_0]$ to minimize

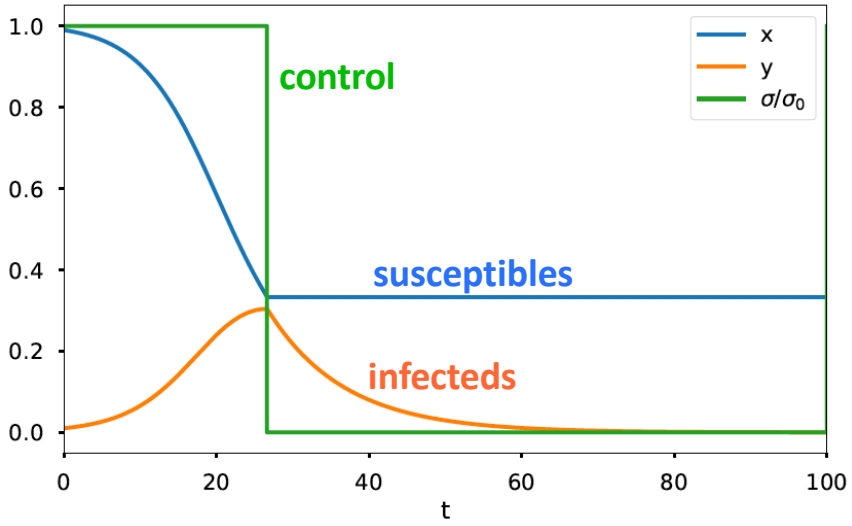
$$J(x(t), y(t), \sigma(t)) = -\lim_{t\to\infty} x(t) + \int_0^T L(x(t), y(t), \sigma(t))dt$$

subject to $x'(t) = -\gamma\sigma(t)yx$

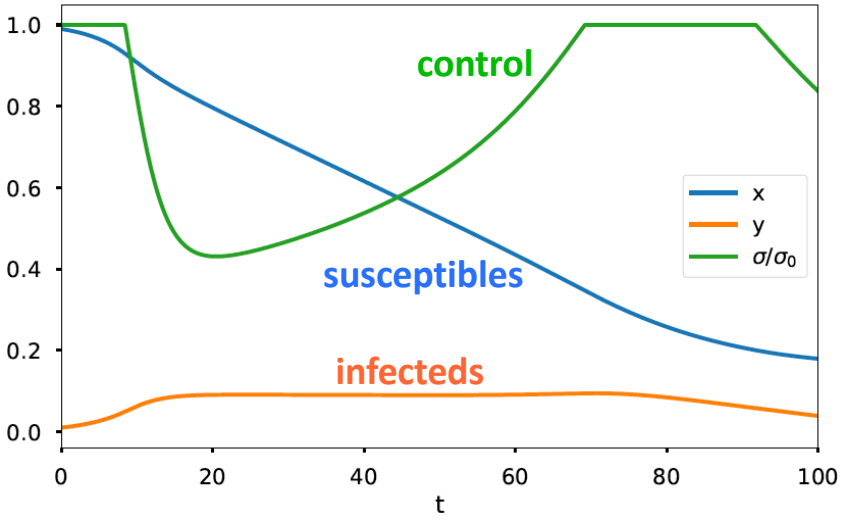
$$y'(t) = \gamma\sigma(t)yx - \gamma y$$

Optimizing Response (D. Ketcheson, KAUST)

Control with no cost



Optimal solution
with constraint on hospital capacity



Journal of Theoretical Biology 511 (2021) 110557

Contents lists available at

Recently, optimal control approach has been introduced to minimize the final size by temporary reduction of the contact rate on a given time interval $[0, D], D > 0$. This issue has been considered in Ketcheson (2020), with total lockdown and added integral term accounting for control cost; and in Bliman et al. (2020), where partial lockdown is considered as well. The corresponding optimal control is bang-bang, with maximal distancing intensity applied

How b

Pierre-Al

[a] Inria, Sorbonne [illegible], Université Paris-Diderot SPC, CNRS Laboratoire Jacques-Louis Lions, équipe MAMBA, Paris, France
[b] Inria, Université de Strasbourg, ICUBE, équipe MIMESIS, Strasbourg, France

Check for updates

Tracking spread (I. Alam, KAUST)

CovMT: an interactive SARS-CoV-2 mutation tracker, with a focus on critical variants

The number of confirmed SARS-CoV-2 cases worldwide has now reached around 100 million, with 2·1 million reported deaths[1] and more than 450 000 SARS-CoV-2 genomes already sequenced. It is vital to keep track of mutations in the genome of SARS-CoV-2, especially in the spike protein's receptor binding domain (RBD) region, which could potentially impact disease severity and treatment strategies.[2–4] In the wake of a recent increase in cases with a more infective variant featuring an RBD mutation (N501Y, B.1.1.7) in the UK, countries worldwide are

CovMT also provides a timeline history of SARS-CoV-2 variants related to mutations in the RBD region of the spike protein. As of the end of January, 2020, the spread of N501Y, B.1.1.7 variants has been detected in SARS-CoV-2 isolate genomes from nearly 60 additional countries using CovMT (appendix). Nonsynonymous mutations in the RBD region have a high potential to be linked to increased binding efficiency, increased infectivity, and the potential to evade antibodies.[2–4] To track all similar variants, we ranked mutations in the RBD region based on their appearance in the number of isolate genomes in CovMT. The CovMT timeline (appendix) shows that N501Y, S477N, N439K, and L452R mutations can now

With a particular focus on critical mutations in the RBD region of the spike protein, and with an option to seamlessly accrue the clinical metadata, including disease severity, we believe that CovMT will be useful for scientists, the general public, and authorities to explore country-specific information.

We declare no competing interests. We are thankful to King Abdullah University of Science and Technology information technology and supercomputing laboratory teams for maintaining the computational resources and helping to publish the CovMT website, and to GISAID for providing daily updates on sequenced isolates worldwide. This work is supported by King Abdulaziz City of Science and Technology grant for COVID-19 research, number 0004-002-01-20-5.

*Intikhab Alam, Aleksandar Radovanovic, Roberto Incitti, Allan A Kamau, Mohammed Alarawi, Esam I Azhar, Takashi Gojobori
intikhab.alam@kaust.edu.sa

Lancet Infect Dis 2021
Published Online February 8, 2021
https://doi.org/10.1016/S1473-3099(21)00078-5

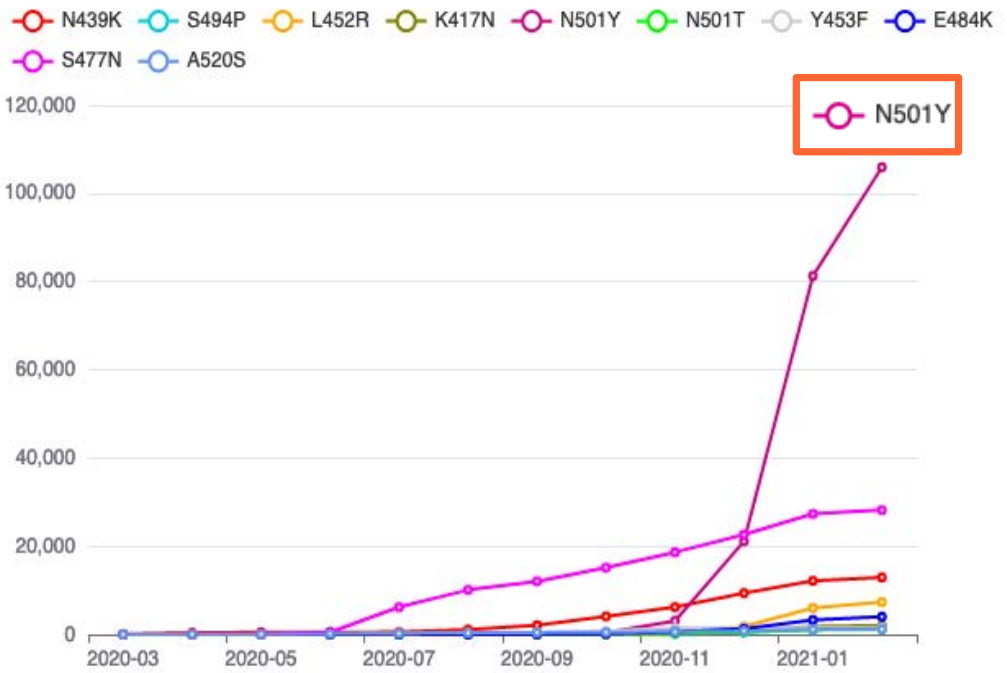
- **114,545,709 cases worldwide with 636,464 genomes sequenced in GISAID* (~0.56%)**
- **Mutations in the RBD** region linked to increased binding efficiency, increased infectivity, and the potential to evade antibodies**
- **Web interface mutation tracking system provided for SARS-CoV-2 variants**

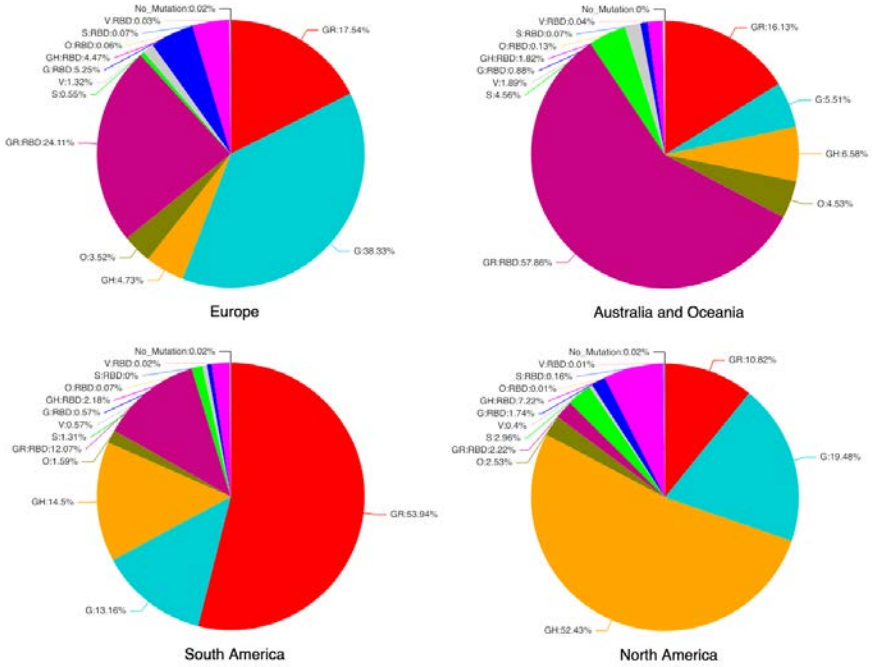
* Global Initiative on Sharing Avian Influenza Data

* Receptor Binding Domain

Tracking spread (I. Alam, KAUST)

https://www.cbrc.kaust.edu.sa/covmt/





"This is helpful for tracking variants that might make vaccines less effective. If these variants become widespread they make eradicating transmission of the virus much more difficult to achieve."

Thomas Christie Williams, University of Edinburgh & COG-UK

Fluid-dynamical transport (V. Vuorinen, Aalto)

Safety Science 130 (2020) 104866

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

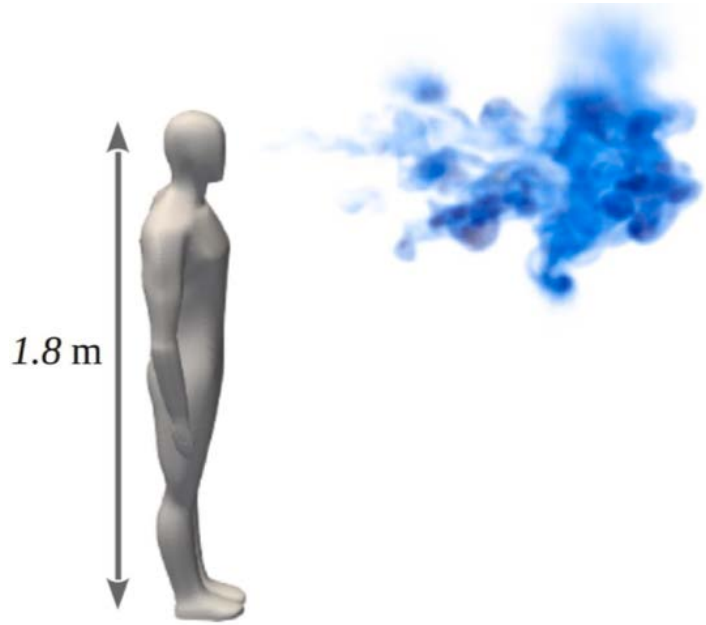
Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors

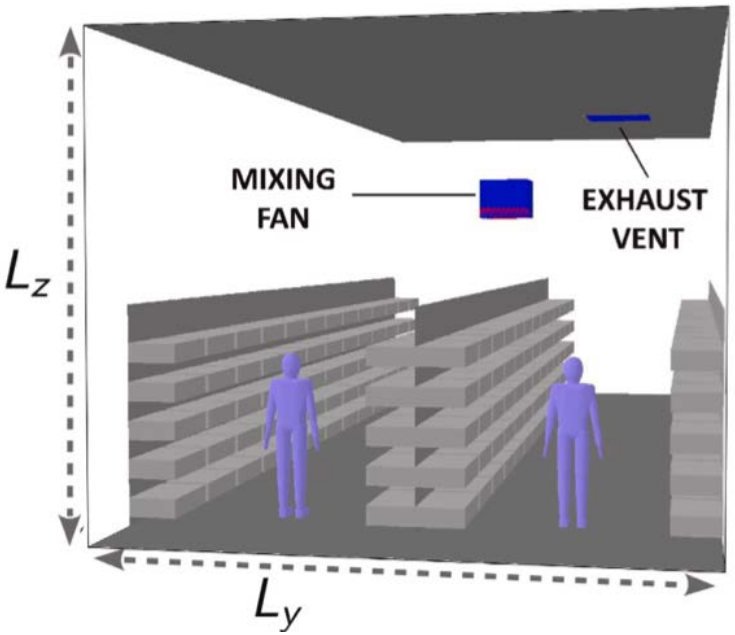
Ville Vuorinen[a,*], Mia Aarnio[b,1], Mikko Alava[h], Ville Alopaeus[c], Nina Atanasova[b,i], Mikko Auvinen[b], Nallannan Balasubramanian[e], Hadi Bordbar[g], Panu Erästö[f], Rafael Grande[d], Nick Hayward[e], Antti Hellsten[b], Simo Hostikka[g], Jyrki Hokkanen[m], Ossi Kaario[a], Aku Karvinen[l], Ilkka Kivistö[l], Marko Korhonen[h], Risto Kosonen[a], Janne Kuusela[n], Sami Lestinen[a], Erkki Laurila[a], Heikki J. Nieminen[e], Petteri Peltonen[a], Juho Pokki[c], Antti Puisto[h], Peter Råback[m], Henri Salmenjoki[h], Tarja Sironen[j,k], Monika Österberg[d]

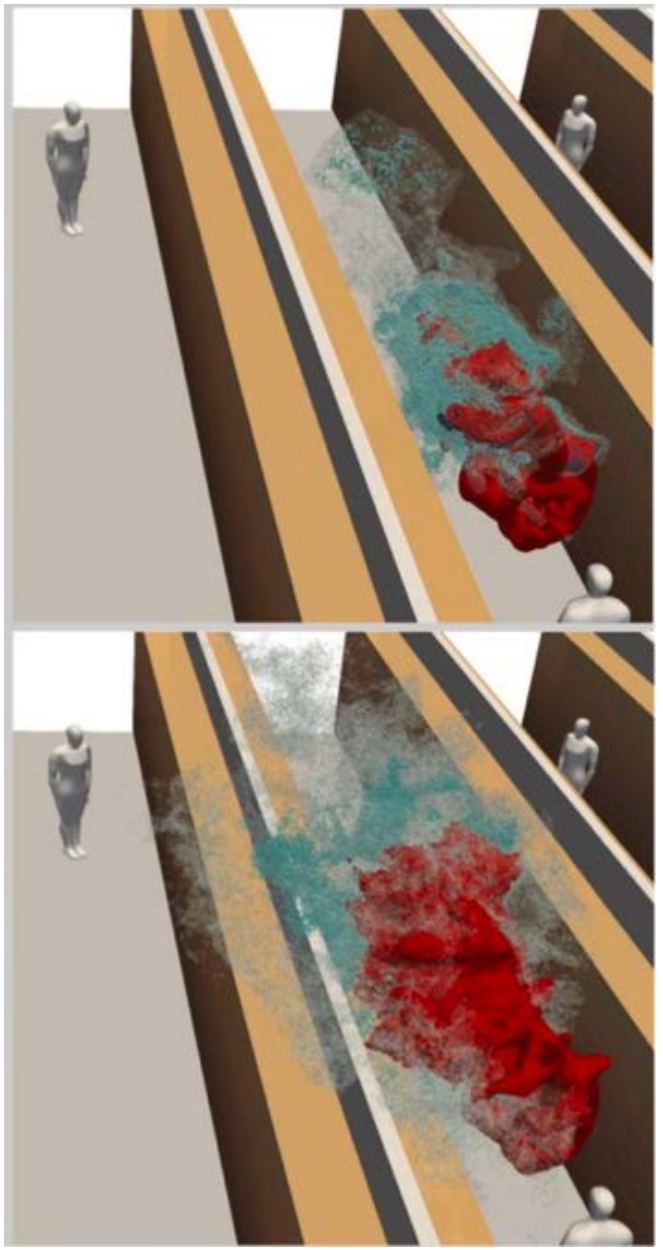
- **Physics of dispersion of aerosol droplets of diameter < 20 µm**
- **Large Eddy Simulation Boussinesq (using PALM*, OpenFOAM*, FDS*, NS3dLab)**
  - **incompressible Navier-Stokes for air velocity + scalar transport eqn *or* discrete Lagrangian stochastic particle model**
  - **find "Domain of Elevated Risk" based on threshold of integrated aerosol intake**

*open source, configurations available from authors

Fluid-dynamical transport (V. Vuorinen, Aalto)







Molecular Dynamics (R. Amaro, UCSD)

AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics

Lorenzo Casalino[1†], Abigail Dommer[1†], Zied Gaieb[1†], Emil P. Barros[1], Terra Sztain[1], Surl-Hee Ahn[1], Anda Trifan[2,3], Alexander Brace[2], Anthony Bogetti[4], Heng Ma[4], Hyungro Lee[5], Matteo Turilli[5], Syma Khalid[6], Lillian Chong[4], Carlos Simmerling[7], David J. Hardy[3], Julio D. C. Maia[3], James C. Phillips[3], Thorsten Kurth[8], Abraham Stern[8], Lei Huang[9], John McCalpin[9], Mahidhar Tatineni[10], Tom Gibbs[8], John E. Stone[3], Shantenu Jha[5], Arvind Ramanathan[2*], Rommie E. Amaro[1*]
[1]University of California San Diego, [2]Argonne National Lab, [3]University of Illinois at Urbana-Champaign, [4]University of Pittsburgh, [5]Rutgers University & Brookhaven National Lab, [6]University of Southampton, [7] Stony Brook University, [8]NVIDIA Corporation, [9]Texas Advanced Computing Center, [10]San Diego Supercomputing Center,
[†]Joint first authors [*]Contact authors: ramaro@ucsd.edu, ramanathana@anl.gov

COVID Gordon Bell Prize winner

ABSTRACT

We develop a generalizable AI-driven workflow that leverages heterogeneous HPC resources to explore the time-dependent dynamics

'20: International Conference for High Performance Computing, Networking, Storage, and Analysis. ACM, New York, NY, USA, 14 pages. https://doi.org/finalDOI

- **Scientific merits: investigation of SARS-CoV-2 spike protein mechanism**
  - **Elludicate invisible mechanisms, including glycan shield and glycan modulating role**
  - **Classical explicit Molecular Dynamics on 305M-atom viral envelope (up to 4.2 μs)**
- **Infrastructure merits: efficiencies applied at new scales on several large systems**
  - **Algorithmic: weighted ensemble strategy in MD – efficiently kill uninteresting trajectories using ML over many short parallel sims parallel, vs. one long serial sim**
  - **Software: wide portability across new architectures (CHARM36, NAMD2.14)**

Molecular Dynamics (R. Amaro, UCSD)

"Computational microscopy"

1. ***Spike protein of* 600K *atoms an order of magnitude larger than previous (100TB per sim)***
2. ***Composite system to view "handshake" has* 8.5M *atoms with explicit water***
3. ***Full protein envelope has* 305M *atoms***

Linked in AI-driven workflow

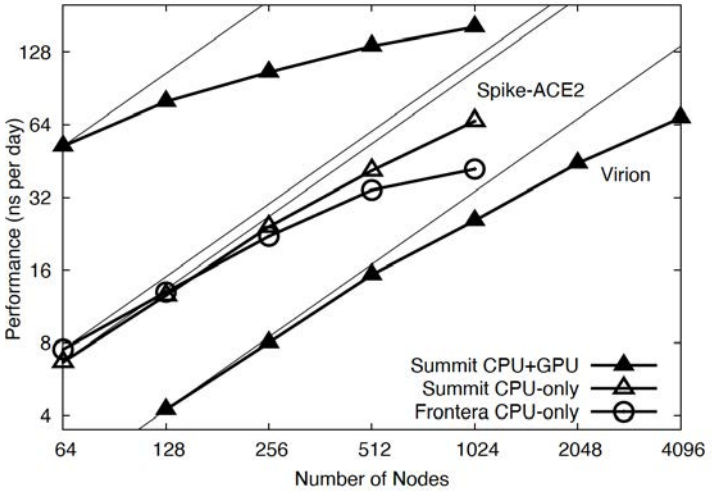
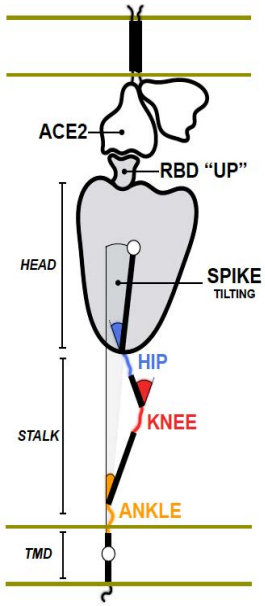
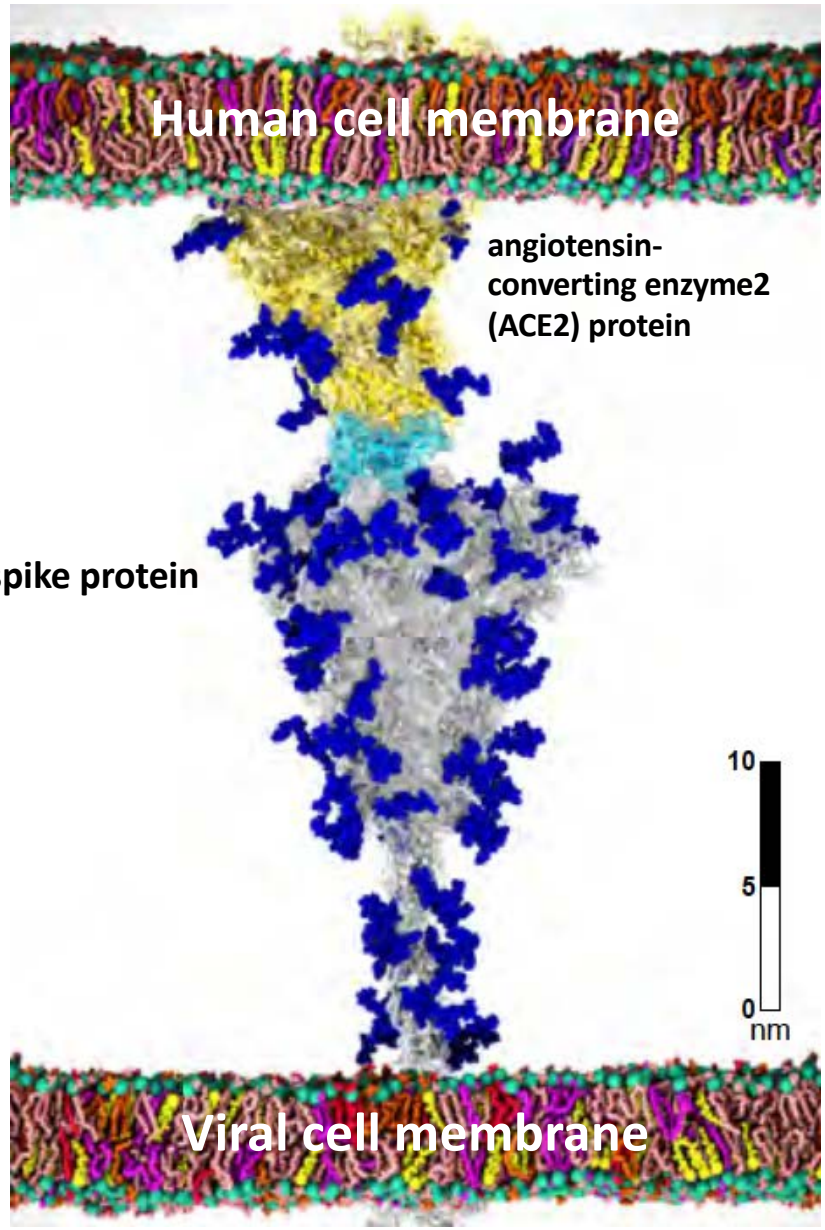


Figure 4: NAMD scaling on Summit and Frontera for 8.5M-atom spike-ACE2 complex (upper lines) and 305M-atom virion (lower line). Thin lines indicate linear scaling.





Drug discovery/repurposing (H. Arthanari, Harvard)

iScience

CellPress
OPEN ACCESS

Gorgulla et al., iScience 24, 102021
February 19, 2021 © 2020 The Author(s).
https://doi.org/10.1016/j.isci.2020.102021

Article

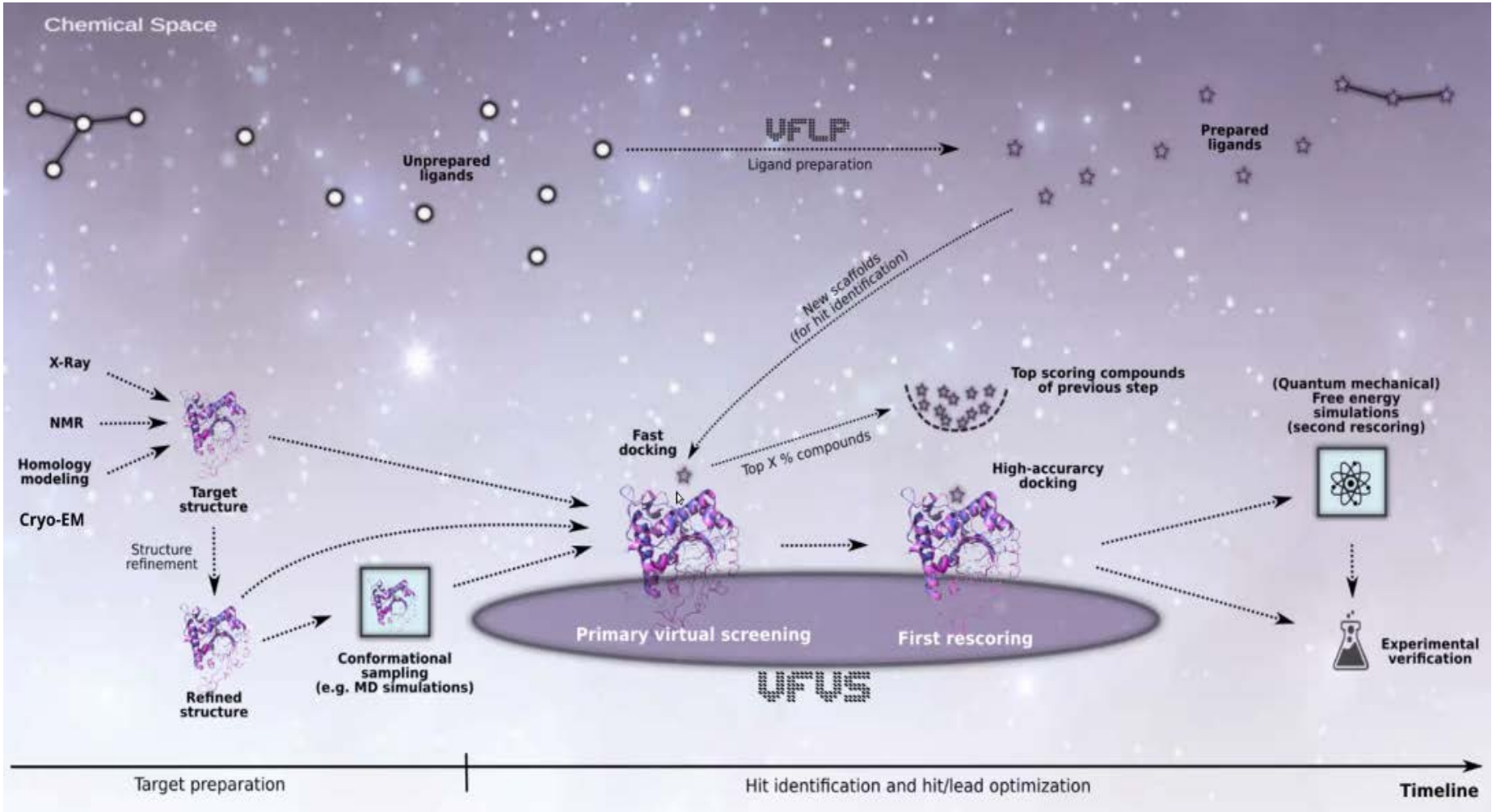
A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening

Christoph Gorgulla,[1,2,3,*] Krishna M. Padmanabha Das,[1,3] Kendra E. Leigh,[11,20,21] Marco Cespugli,[13] Patrick D. Fischer,[1,3,9] Zi-Fu Wang,[1] Guilhem Tesseyre,[4] Shreya Pandita,[4] Alec Shnapir,[4] Anthony Calderaio,[17] Minko Gechev,[4] Alexander Rose,[16] Noam Lewis,[4] Colin Hutcheson,[4] Erez Yaffe,[4] Roni Luxenburg,[4] Henry D. Herce,[1,3] Vedat Durmaz,[13] Thanos D. Halazonetis,[12] Konstantin Fackeldey,[8,10] J.J. Patten,[15] Alexander Chuprina,[5] Igor Dziuba,[19] Alla Plekhova,[6] Yurii Moroz,[6,7] Dmytro Radchenko,[5,7] Olga Tarkhanova,[6] Irina Yavnyuk,[5] Christian Gruber,[13,14] Ryan Yust,[4] Dave Payne,[4] Anders M. Näär,[18] Mark N. Namchuk,[1] Robert A. Davey,[15] Gerhard Wagner,[1] Jamie Kinney,[4] and Haribabu Arthanari[1,3,22,*]

- **VirtualFlow used to search for SARS-CoV-2 inhibitors**
  - **1 billion small molecules screened *in silico***
  - **40 different docking sites**
  - **17 different potential viral and host targets**
- **In addition to targeting the active sites of viral enzymes, targeted critical auxiliary sites such as functionally important protein-protein interactions**

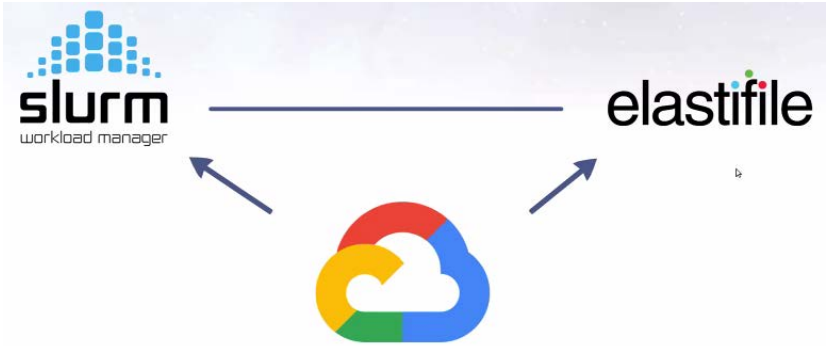
Drug discovery/repurposing (H. Arthanari, Harvard)

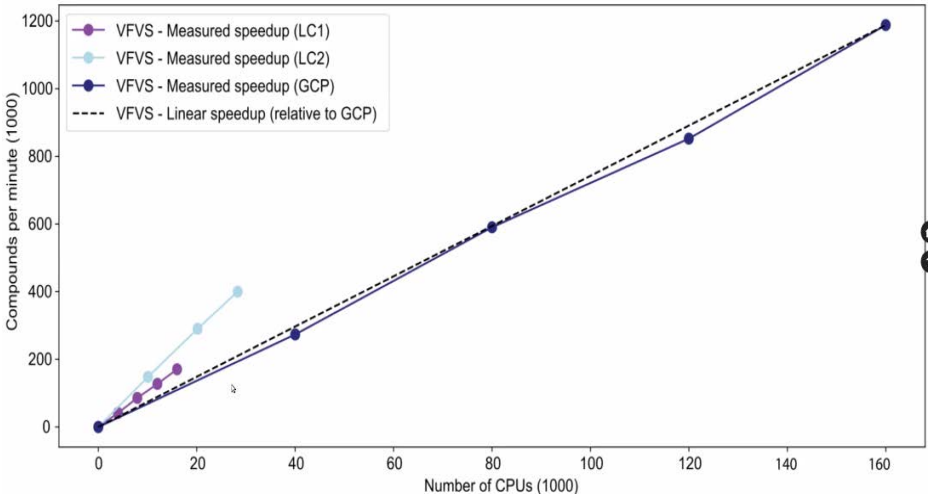
Workflow: initial coarse and secondary refined screening of promising hits



Drug discovery/repurposing (H. Arthanari, Harvard)

- **In addition active sites of viral enzymes, critical auxiliary sites such as functionally important protein-protein interactions are targeted**
- **50 billion docking instances**
- **100M CPU hours in the Google cloud**





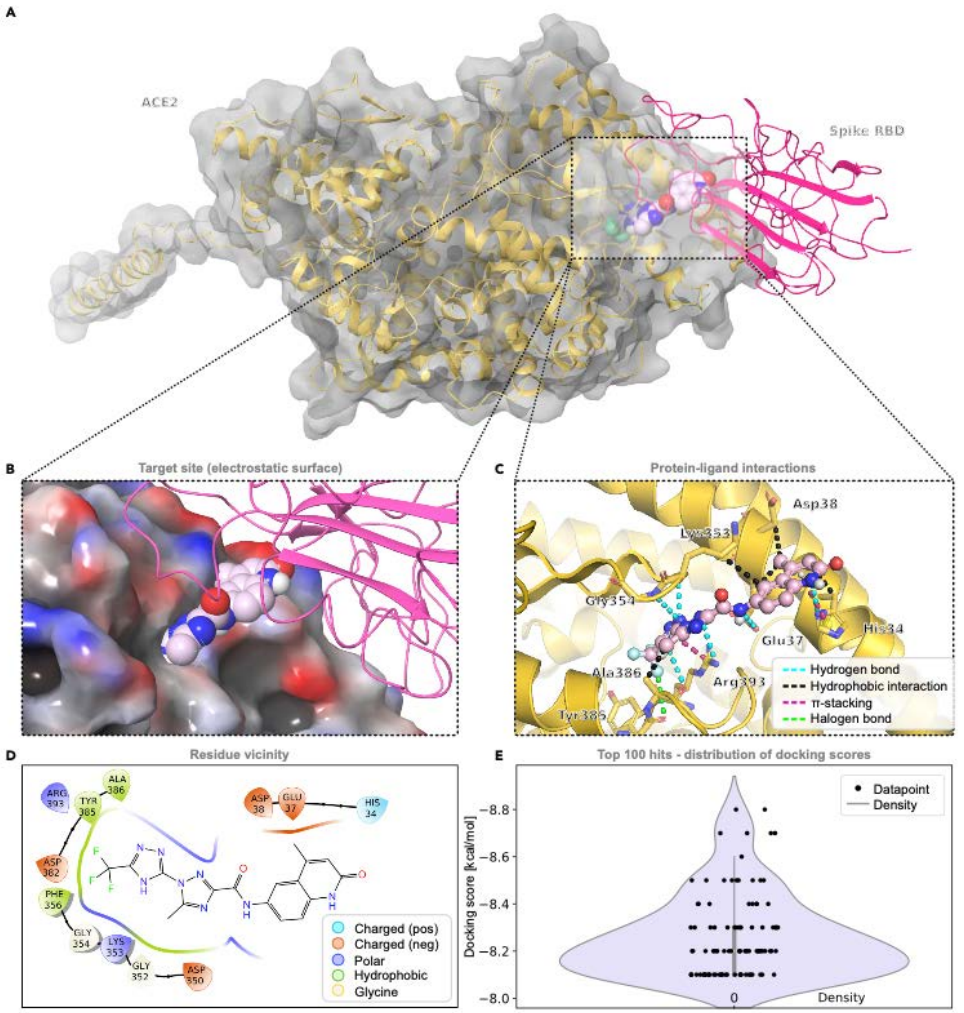


Figure 2. The ACE2 receptor and an example compound from the top 0.0001% of screened compounds bound at the spike interaction interface (site1)

(A) The target protein ACE2 (gold) bound to the RBD of the spike protein (magenta) and an example compound (light pink) from the virtual screen bound to the spike interaction interface (site 1, around Glu37) (screen ID: 1).
(B) The electrostatic surface of the target protein (ACE2) bound to the RBD domain of the spike protein (magenta) and an example compound (light pink).
(C) An overview of the interactions between the ligand and the receptor structure.
(D) Receptor residues within 4 Å of the ligand.
(E) Distribution of the docking scores of the top 100 virtual screening hits.

Drug discovery/repurposing (H. Arthanari, Harvard)

https://vf4covid19.hms.harvard.edu/

AI is a vital part of the mix

- Allowed MD modelers to “fast forward” over uninteresting stretches of dynamics and to stitch together models at different scales
- Allowed expensive simulation kernels to be replaced with inexpensive input-output functional maps, leading to improved throughput
- Improved on human accuracy, shrinking both false positives and false negatives in clinical tests
- Allowed “most likely” drug candidates to be identified out of billions of compounds after being trained on a small sample of the compounds for their ability to dock on human cell receptors

Summary (1): CSE has been good for COVID19

- Allowed scientists to visualize billion-atom biomolecular clusters in far greater detail than any laboratory imaging system
- Allowed scientists to understand the ACE2 “attack” configuration of the spike protein with its glycan “kickstand” supports
- Allowed health & safety experts to understand convective and diffusive transport of the virus in rooms and passenger vehicles
- Gave clinicians access to new AI-assisted CT scan interpretation
- Identified hundreds of potential new drugs computationally and launched dozens into experimental pipelines and clinical trials
- Created computational infrastructure and workflows relevant to other pandemics
- Enabled public policy makers to smartly dial the balance between losing lives and suspending livelihoods

Summary (2): COVID19 has been good for CSE

- Pushed out #1-ranked system a year ahead of schedule
- Created new continental consortia of HPC centers
- Led to sharing of data on unprecedented scale
- Created new computer-to-laboratory pipelines
- Focused new attention on the role of preprint archives
- Accelerated injection of AI into scientific and clinical workflows
- Accelerated development of new algorithms in MD
- Accelerated development of new stratified models in epidemiology
- Brought billions of readers to the precipice of computational modeling and machine learning, “looking in” from popular media accounts and inspiring future scientists

Summary (3): COVID19 has been good for CSE ☺

"The pandemic hibernation has made the engine lab useless and thus many of my students have converted to simulations. That has more than doubled the use of high-CPU simulations. I do not see this change back to more experiments any time soon..."

– a combustion colleague

Computational Projects on KAUST's Cray XC-40

PI	Institution	Project Title
Alzeer	KSA MoH & Harvard SPH	KSA Dynamic Transmission Microsimulation Model (DTM2)
AbdulAzeez	Iman Abdulrahman U	*In silico* Screening, Molecular Docking and Dynamic Simulations of Drugs for Treating COVID-19
Aldulajian	Iman Abdulrahman U	Molecular Modeling-Based Design of Peptides to Inhibit Infection by SARS-CoV2
Ibrahim	Jouf U	SARS-CoV-2-RNA Dependent RNA Polymerase Targeting
Shabbir	King Khaled U	The Study of Potential Inhibitor Candidates for COVID-19 Main Protease Among Medicinally Bioactive Compounds: *in silico* Computational Screening
Imran	Northern Border U	Network Analysis-based Characterization of Novel COVID-19 Molecular Pharmacophores Targeting Acute Viral Infection *in vitro* Mechanism
Baz	Taif U	Prediction Model Covid19 Outbreak in KSA based on AI
Cavallo	KAUST	SARS-CoV-2: A Molecular Modelling and Simulation Approach
Gao	KAUST	Elucidating the Functioning Mechanism of Remdesivir Targeting RdRp and Repurposing Drugs for COVID-19
Hauser	KAUST	Molecular Dynamic Simulations to Study Binding Interactions of Small Molecules with Epitopes of the SARS-CoV-2 Spike Protein
Pain	KAUST	Comparative Genomics Study of SARS-CoV-2 Genome
Keyes	KAUST	Unstructured Mesh Deformation Using Rank Structured Matrix Computations: a Preliminary Study with the SARS-CoV-2 Virus

Computational Projects on KAUST's Cray XC-40

PI	Institution	Code(s)	Project Description	Usage
Alzeer	KSA MoH & Harvard SPH	Harvard SPH code	agent-based SEIR model, parameter fit to MoH case data, for forward projection	throughput
AbdulAzeez	Iman Abdulrahman U	GROMACS, AutoDock, NAMD	*in silico* screening of existing drugs on various viral proteins	throughput
Aldulajian	Iman Abdulrahman U	GROMACS, NAMD, CP2k	*in silico* screening of bioactive peptides to inhibit binding to epithelial cells	throughput
Ibrahim	Jouf U	NAMD, VMD	investigate effect of mutations of the RNA polymerase on antiviral drugs	throughput
Shabbir	King Khaled U	GAMESS, AutoDock	*in silico* screening of existing drugs to inhibit the 3CL Mpro protease that cleaves the viral RNA	throughput
Imran	Northern Border U	GROMACS, CHARMM	*in silico* screening of existing drugs targeting hydrophobic interactions	throughput
Baz	Taif U	In-house Python	short-term case prediction using regression	throughput
Cavallo	KAUST	GROMACS, AutoDock	*in silico* screening of drugs to inhibit the dimerization of Nsp9 (RNA-binding protein)	throughput
Gao	KAUST	GROMACS	all-atom classical MD investigating Remdesivir impact on addition & cleavage sites of nsp12 & nsp14	throughput
Hauser	KAUST	GROMACS	*in silico* screening of small molecules with spike protein binding sites	throughput
Pain	KAUST	RAxML, MAFFT	investigate nucleotide variation in published viral genomes	throughput
Keyes	KAUST	HiCMA	mesh generation and deformation for transport of viruses in fluids, using RBFs [Mesri's talk in MS12]	capability

Thank you!
جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology
شكرا
david.keyes@kaust.edu.sa

Where to find these slides

https://cemse.kaust.edu.sa/sites/default/files/ms78keyes.pdf

[See

https://cemse.kaust.edu.sa/ecrc/news/ecrc-contributions-siam-cse21/

for **all 17** ECRC contributions to CSE 2021]