UNICAMP

Aeronautical Sciences in the Era of Big Data

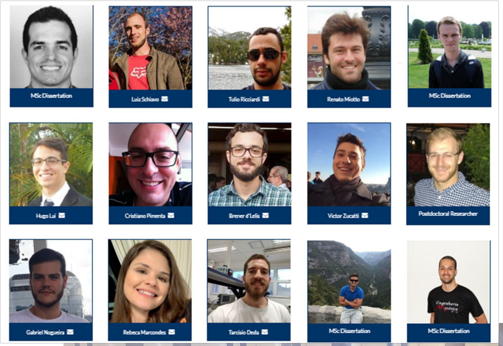
William R. Wolf
wolf@fem.unicamp.br

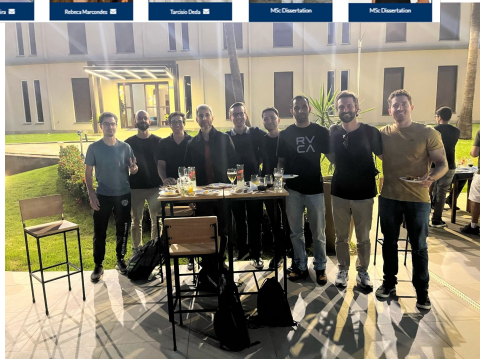
School of Mechanical Engineering
University of Campinas

Love Data Day, June 2025

Laboratory of Aeronautical Sciences

PI: William Wolf





- Research interests: unsteady aerodynamics, turbulence, aeroacoustics, hydrodynamic instabilities, flow control

- Approach: high-fidelity simulations, reduced-order modeling, data science

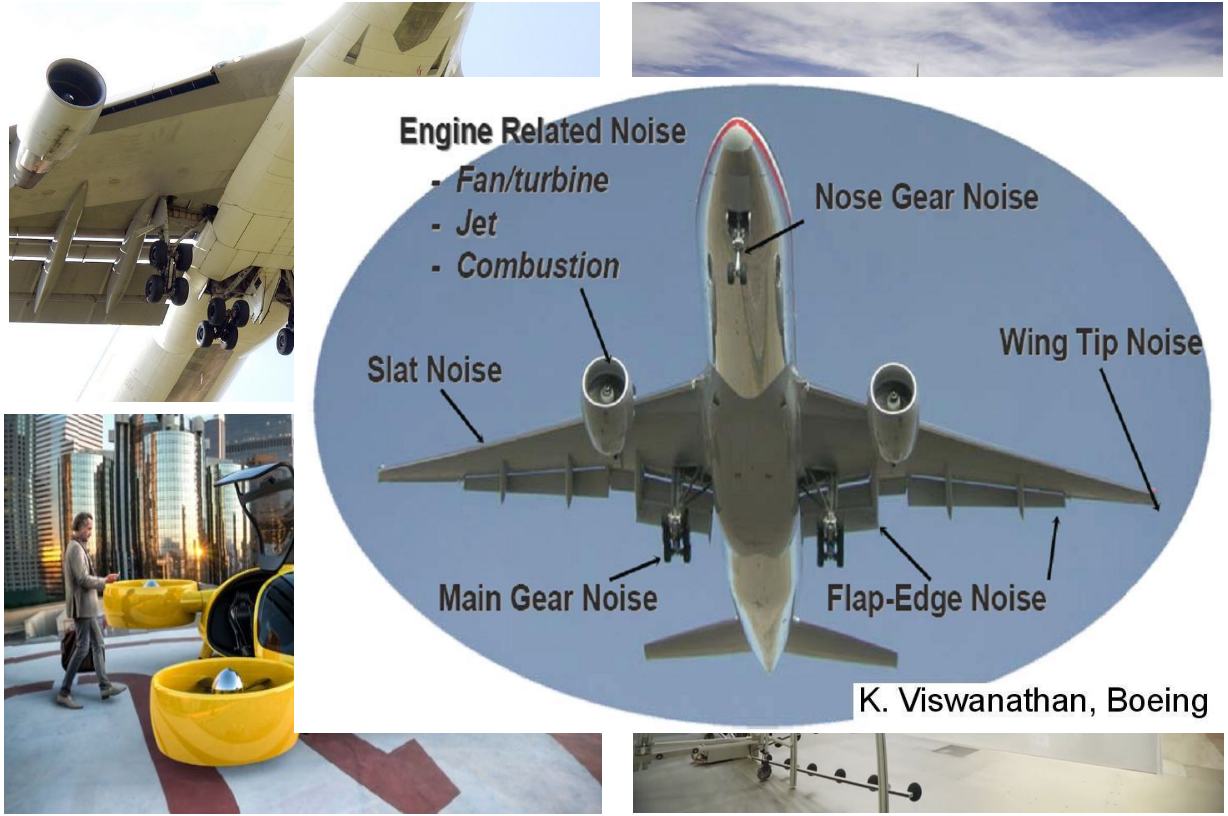
Turbulence and turbulent flows

- Turbulence: “The most important unsolved problem of classical physics” (Feynman)
- Nature of turbulence: irregularity, mixing, three-dimensional, eddies
- Applications in nature, industry and everyday life: transportation and energy, weather, industrial processes
- Advances in computational power and numerical methods allow the accurate prediction of turbulent flows

Aeroacoustics: noise from aerodynamic flows

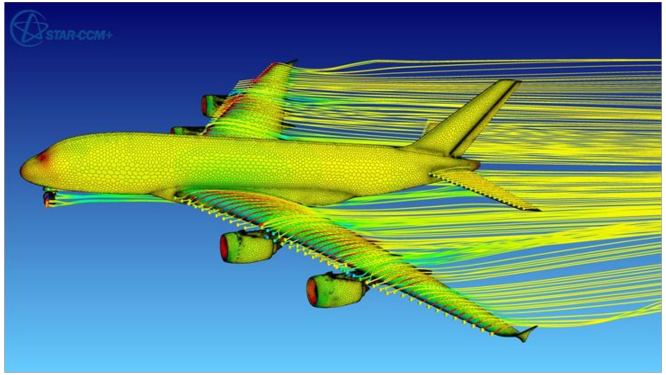
- Aeroacoustics: field of study that deals with sound generation and propagation in unsteady (turbulent) aerodynamic flows
- Applications: aircraft, automobiles, wind and gas turbines, musical instruments, fans and rotors, home appliances

Aeroacoustics: noise from aerodynamic flows



- Aeroacoustics: field of study that deals with sound generation and propagation in unsteady (turbulent) aerodynamic flows
- Applications: aircraft, automobiles, wind and gas turbines, musical instruments, fans and rotors, home appliances

Computers and the aeronautical industry



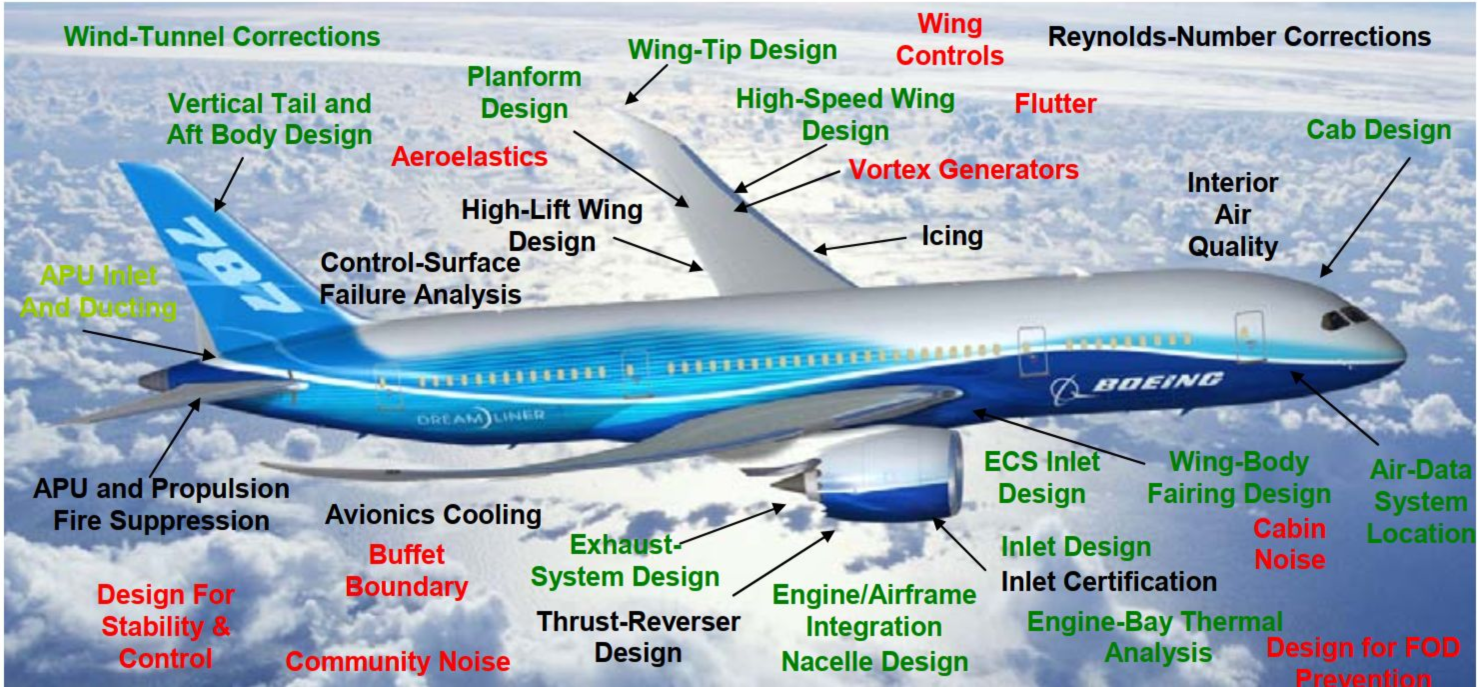
- In the past, industry relied on expensive wind tunnel and flight tests
- Computers allowed design via CFD (Computational Fluid Dynamics)
- 1970’s: first simulations of wings (low fidelity)
- 1980’s: first supercomputers allowed simulations of full aircraft (still low fidelity)
- 1990’s: CFD becomes commercial success
- 2000’s: high-fidelity simulations become possible

Role of CFD on design of modern aircraft

Extensive CFD Contributions

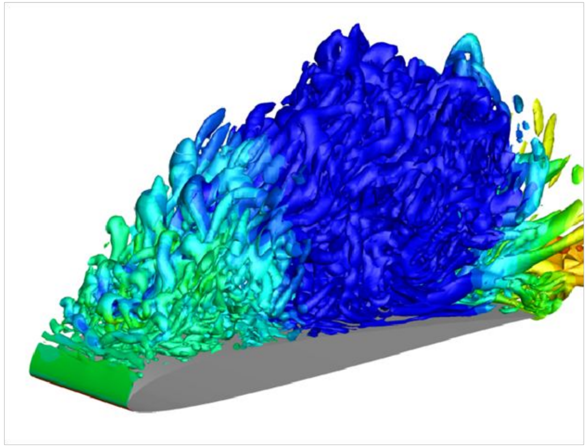
Moderate CFD Contributions

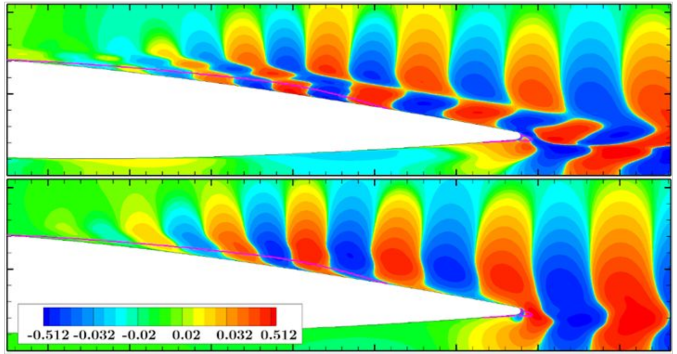
Some CFD Contributions



Tinoco, Boeing (2007)

What do we do in the laboratory?





- **High-fidelity simulations:** formulations based on first principles, avoid modeling, numerical algorithms that mimic the physics, use of **intense computational power**, generation of **large datasets**

- **Reduced-order models:** appropriate for design, optimization and flow control, lower computational power, should represent the physics with higher fidelity possible, **need for data**

Coaraci (CEPID-CCES) e Lovelace (CENAPAD-SP)

Novo supercomputador é ativado na Unicamp

Chegada da "supermáquina" Lovelace é festejada no Cenapad-SP, que tem sede na Unicamp

| Autoria Luiz Sugimoto | Fotos Acervo Cenapad | Antonio Scarpinetti | Edição de imagem Renan Garcia

Turbulence is BIG data

UNICAMP

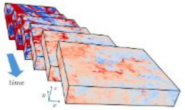
Home | Database Access | Datasets | Citing | Publications | Visualizations

NOTICE: Four new datasets (from a 32,768-cubed record-breaking DNS of isotropic turbulence, a strongly stable LES of atmospheric boundary layer, and two wind farm flows) are now available. Also, the data access methods now include Python and Matlab, Fortran and C. If you are still using the old methods, please consider updating to the new ones (essentially replacing "getVelocity(..)", etc. calls with the new "getData(..)" function).

NOTICE: All services are functioning normally. For past announcements, please click here

649.798.164.903.808 points queried

Welcome to the Johns Hopkins Turbulence Database JHTDB



This website is a portal to an **Open Numerical Turbulence Laboratory** that enables access to multi-Terabyte turbulence databases. The data reside on distributed nodes and disks on our database cluster computers and are stored in small 3D sub-cubes to accelerate access speeds to small data subsets.

Access to the data is facilitated by **Web Service Interfaces** that permit numerical experiments to be run across the Internet. We offer Python, Matlab™, Fortran and C interfaces layered above Web Services so that scientists can use familiar programming tools. Python, Matlab™, Fortran and C codes can be run on client platforms. Python can also be run on Sciserver, a fully integrated cyberinfrastructure system encompassing related tools and services to enable researchers to cope with scientific big data. Subsets of the data can be downloaded in hdf5 file format using the data cutout service. For initial experimentation and familiarization of datasets and access modes, web-browser manual querying at individual points and times is also supported.

Data processing supporting queries for velocity, pressure, gradients, Laplacians, and Hessians (and other relevant variables depending on the dataset) at arbitrary points and time is supported using Lagrange Polynomial and Spline methods executed on the database close to the data. For time-resolved datasets, fluid particle tracking can be performed both forward and backward in time using a second order accurate Runge-Kutta integration scheme.

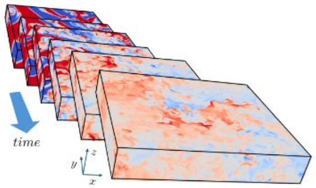
Available **datasets** (see datasets description page) include a **space-time history** of a direct numerical simulation (DNS) of isotropic turbulence (100 Terabytes), a DNS of magneto-hydrodynamic (MHD) turbulence (50 TB), a DNS of forced, fully developed turbulent channel flow (130 TB), a DNS of homogeneous buoyancy driven turbulence (27 TB), a transitional boundary layer flow (105 TB), data from Large Eddy Simulations (LES) of stably stratified atmospheric turbulent boundary layer (40 TB) and two wind farm LES: one in a conventionally neutral flow and another during a diurnal cycle (JHTDB-wind datasets generated in collaboration with the Ralph O'Connor Sustainable Energy Institute).

Also available are **datasets** comprising **snapshots** (spatially but not temporally resolved data) of 4096^3 DNS of isotropic turbulence (1 snapshot), 8192^3 DNS of isotropic turbulence (6 snapshots at higher Reynolds number), a single snapshot of a record-breaking $32{,}768^3$ DNS of isotropic turbulence at R_λ ~ 2,550 (1/2 petabytes), rotating stratified turbulence (5 snapshots, 5 Terabytes), a high Reynolds number channel flow (11 snapshots, 20 Terabytes) and 40 snapshots of large-eddy simulation data for stably stratified atmospheric turbulent boundary layer flow.

Recent datasets are being stored using the Zarr storage format with 64-cube "buckets", instead of the Z-curve ordering used for the legacy datasets. Also, data are being migrated to Ceph-FS storage from the original SQL-database and FileDB storage modalities. New more modular Python, Matlab™, Fortran and C codes use the REST interface instead of the legacy SOAP interface.

Turbulence is BIG data

JHTDB



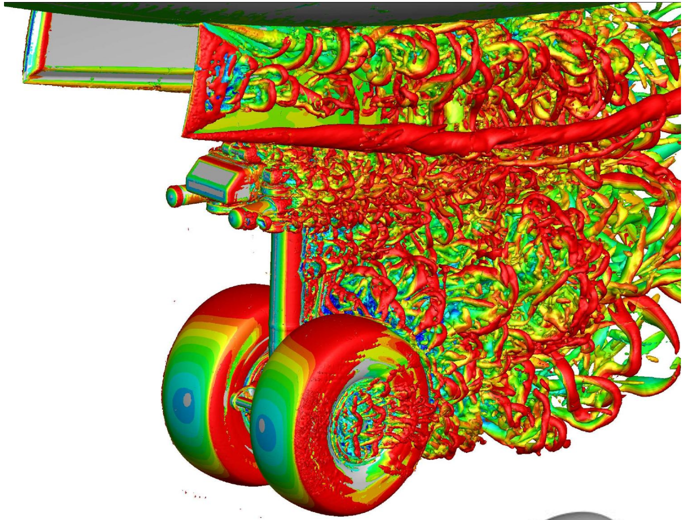
experimentation and familiarization of datasets and access modes, web-browser manual querying at individual points and times is also supported.

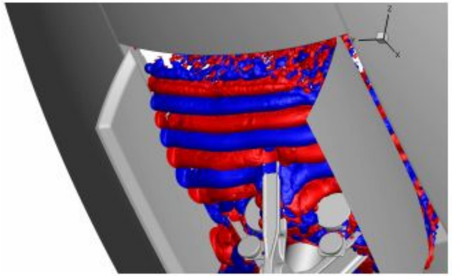
Data processing supporting queries for velocity, pressure, gradients, Laplacians, and Hessians (and other relevant variables depending on the dataset) at arbitrary points and time is supported using Lagrange Polynomial and Spline methods executed on the database close to the data. For time-resolved datasets, fluid particle tracking can be performed both forward and backward in time using a second order accurate Runge-Kutta integration scheme.

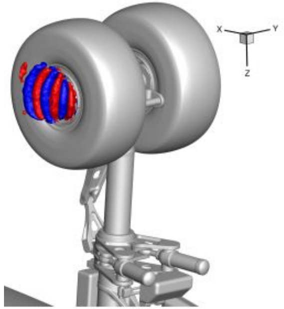
Available **datasets** (see datasets description page) include a **space-time history** of a direct numerical simulation (DNS) of isotropic turbulence (100 Terabytes), a DNS of magneto-hydrodynamic (MHD) turbulence (50 TB), a DNS of forced, fully developed turbulent channel flow (130 TB), a DNS of homogeneous buoyancy driven turbulence (27 TB), a transitional boundary layer flow (105 TB), data from Large Eddy Simulations (LES) of stably stratified atmospheric turbulent boundary layer (40 TB) and two wind farm LES: one in a conventionally neutral flow and another during a diurnal cycle (JHTDB-wind datasets generated in collaboration with the Ralph O'Connor Sustainable Energy Institute).

Noise generation by airframe

Ricciardi et al, JSV 2021

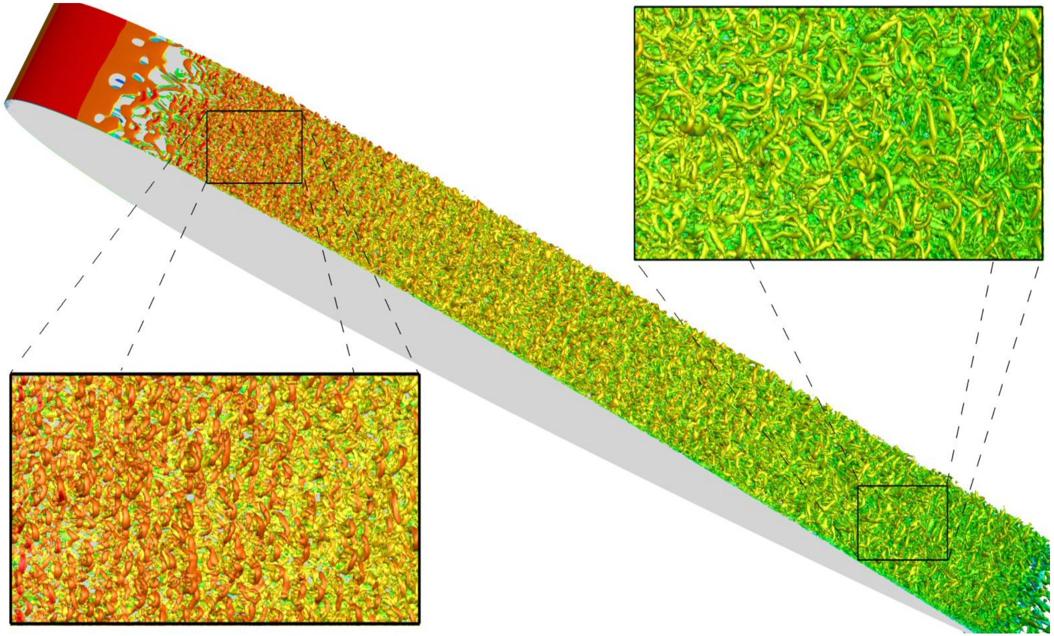






- Design of modern aircraft depends on noise reduction to meet more stringent regulations
- Landing gear becomes important noise source on landing
- Data processing for investigation of noise sources
- Simulation cost: 8M hours!
- **100+ Terabytes of data**

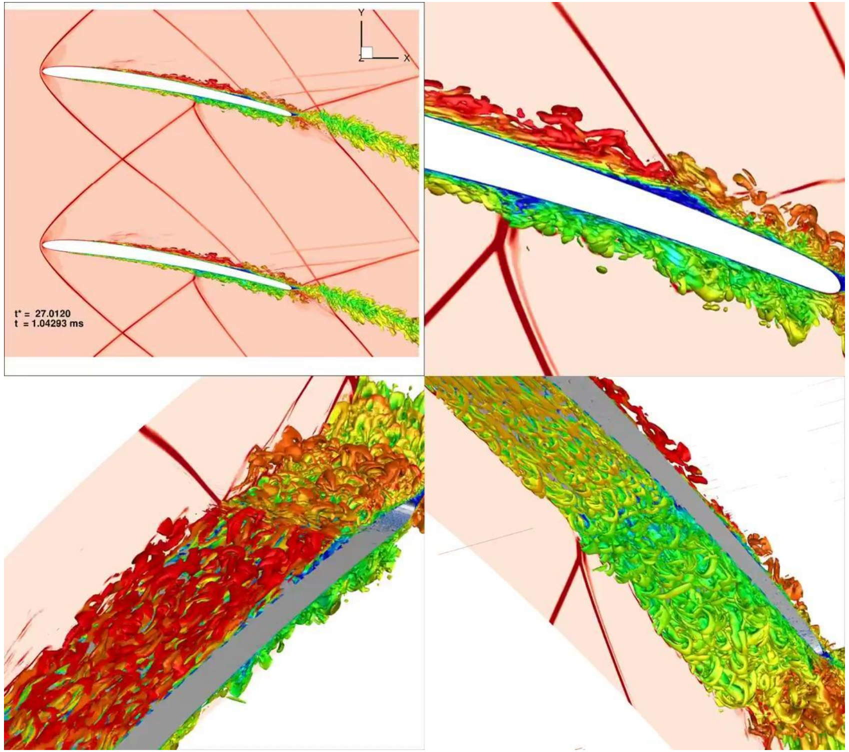
Turbulence generated by wings



Silva and Wolf, IJHFF 2024

- Characterization of turbulence over turbine blades and wings
- Development of flow control strategies for drag and noise reduction: combination of data science and high-fidelity simulations
- Datasets analyzed for years: turbulence relies on statistics
- **80+ Terabytes of data**

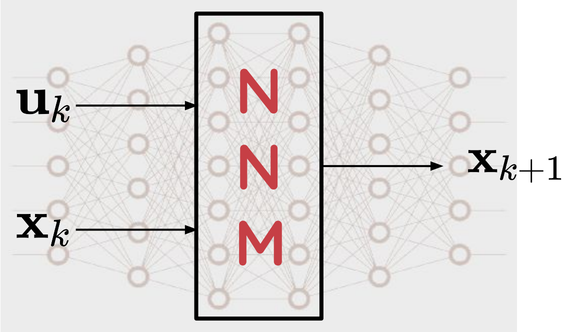
Novel concepts of supersonic turbines

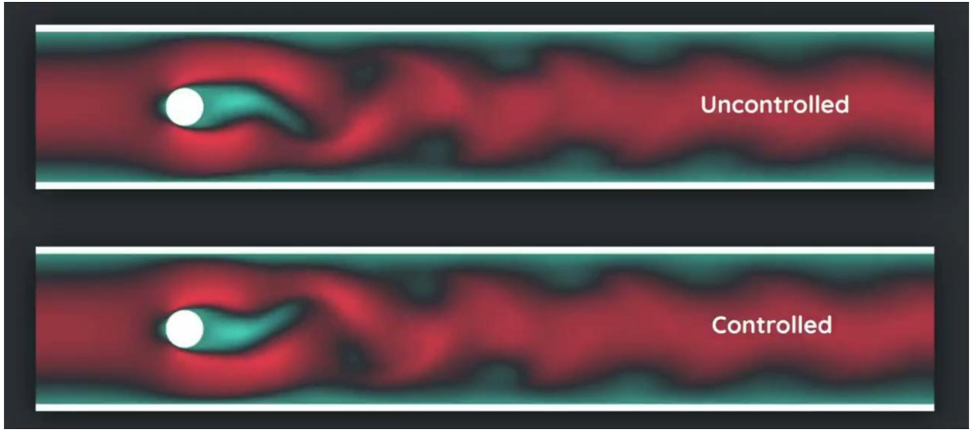


- Applications in high-speed propulsion and energy generation
- Study of the flow physics of shock-waves and boundary layers
- International collaboration with universities in USA, France and Australia (many international groups interested in our data)
- **80+ Terabytes of data**

Lui et al, TCFD 2024

Data science with big data





- Applications of techniques of machine learning to develop:
    - reduced order models
    - real time estimators
    - control modules
    - optimization tools
- Next steps: data used to train real-time experiments
- **Models need data for training, testing and validation!**

UNICAMP

Acknowledgements

CCES UNICAMP

Thank you!

Laboratório Nacional de Computação Científica

XSEDE
Extreme Science and Engineering Discovery Environment

AFOSR
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

FAPESP

CENAPAD SP

BOEING

PRP pró-reitoria de pesquisa unicamp

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

U.S.ARMY

UNICAMP

Aeronautical Sciences in the Era of Big Data

William R. Wolf

wolf@fem.unicamp.br

School of Mechanical Engineering
University of Campinas

Love Data Day, June 2025