

# Tom S. Bertalan

APPLIED SCIENTIST IN DIFFERENTIABLE COMPUTING, PROBABILISTIC INFERENCE, AND NEUROSymbolic PROGRAM SYNTHESIS; APPLIED TO COMPLEX SCIENTIFIC AND ENGINEERING SYSTEMS.

+1 (256) 613-3760 | Tom@TomBertalan.com | TomSB.net | TSBertalan | tomsbdotnet | Tom-Bertalan-00764640 | US Citizen

## EDUCATION

- Princeton University** 2012 - 2018  
*Ph.D. & Masters, Chemical and Biological Engr.*
- The University of Alabama** 2008 - 2012  
*Bachelor of Science in Chemical and Biological Engr.; Minor in Math*

## PROFESSIONAL EXPERIENCE AND RESEARCH AREAS

### Stealth Neurosymbolic AI Startup

FOUNDER

11/2025 - Present

- Prototype an automated scientific model-discovery system that combines symbolic representations, physical constraints, and data-driven components.
- Design and implement a neurosymbolic ML architecture integrating LLM-based semantic reasoning with differentiable physics simulations, and develop hybrid learning pipelines that improve generalization under physical constraints.
- Build experimental pipelines in Python/PyTorch, including Neural ODEs, constrained optimization, and differentiable programming components.
- Lead customer discovery, competitive analysis, and investor/advisor outreach to refine product positioning, MVP roadmap, and pilot strategy.
- Keywords:** Agentic Workflows, Differentiable Computing, Modelica, Neurosymbolic AI, Physics-Informed ML (PIML/SciML), Program Synthesis.

### Amgen

Cambridge, MA

DATA SCIENTIST

11/2024 - 11/2025

- Mechanistic Modeling of Lyophilization and Filling for Biomanufacturing**
  - Designed and advocated a real-time lyophilization humidity digital twin using sequential Bayesian state estimation.
  - Conceived, implemented, and released company-wide a data-validated mechanistic simulation of lyophilization.
  - Calibrated model parameters using experimental data from seven production programs and designed a modular parameterization strategy to support pooled multi-product calibration workflows.
  - Contributed to a growing first-principles mechanistic model of peristaltic filling, used by a broad internal customer base.
  - Implemented data lake ingest pipelines for accelerometer and flow sensors into a new data product for the semantic data fabric.
  - Specified data schemas for a clean library interface, supporting connected and interoperable digital twin architectures across bioprocess operations, and rapid GUI generation on EC2-backed Plotly Dash platform.
  - Led agile sprint for demos for stakeholder requirement capture, and executed three iterative deliveries company-wide of containerized and versioned package releases.
- Hybrid Mechanistic+Black-box Modeling of CHO Cell Metabolism**
  - Evaluated vendor offerings for Gaussian process-based time series system identification tools.
  - Consulted on efficient adjoint gradient-based training of neural differential equations with embedded optimality constraints.
- Agile Development Automation Initiative**
  - Ingested handwritten notes using a multimodal large-language model on Databricks.
  - Accelerated sprint planning using the GitLab and Microsoft Graph APIs.

### University of Massachusetts

Lowell, MA

RESEARCH SOFTWARE ENGINEER

1/2024-11/2024

- Neurosymbolic System Identification of Dynamical Systems**
  - Developed a neurosymbolic framework combining grammatical evolution for modular symbolic program synthesis with neural differential equations, enabling structured reasoning over scientific systems.
  - Ensured cross-domain modularity between robotics and bioprocessing through a domain-specific language design in ANTLR.
  - Prototyped an MVP for LLM-assisted symbolic code generation of differentiable physical simulators.
- Model Predictive Control of CHO Bioreactor**
  - Collected targeted experimental data from a medium-scale bioreactor and developed an online data-driven model-predictive controller with Aspen DMC3 to maximize economic factors, deploying to actual wet-lab hardware.
  - Supervised a rigorous implementation of a SBML-encoded bioreactor digital twin, including a sequential-Bayes soft-sensor for latent bioreactor glucose concentration.
  - Secured grant as PI to develop a central data lake for an NIH-seeded academic/industry consortium.

- **Neural Differential Equations and Time-Series**

- Created a GUI for Bayesian experimental design; mentored team members on its use and maintenance.
- Improved accuracy of backpropagation and convergence time of PyTorch neural stochastic differential equation training.
- Advanced differentiable computing for learning continuous scientific laws from time-series data, bridging symbolic and statistical modeling through neural ODEs, PDEs, and SDEs; resulting in 9 publications and 10 conference presentations.
- Cut RNN inference burn-in from 25 to 5 samples using manifold learning.
- Led a team of biophysics and ML experts in creating a suite of simulation tools for CHO metabolism, including custom gradients for nets with constraints.

- **Robotic Perception and Control**

- Developed a variational autoencoder for end-to-end robotic localization.
- Trained a U-net on both open and custom datasets for real-time (>10hz) on-board semantic segmentation of drivable space.

**The Massachusetts Institute of Technology**

POSTDOCTORAL ASSOCIATE

Cambridge, MA

3/2018-2/2020

- **Nonlinear dynamics in neuroscience and bioprocessing**

- Wrote object-oriented library for fine- and coarse-grained simulations of neuronal dynamics, probing bifurcation and resonance behavior of a mammalian circadian rhythm model.
- Coauthored successful application for a \$1.8M NIH-led industry-academic partnership grant for a new bioreactor simulation.

- **Autonomous Vehicle Design and Pathfinding**

- Developed a model AV with firmware-level speed sensing and control commanded by checksummed bus communications.
- Simulated a comfort-optimized (jerk-minimizing) path planner with two-lane-ahead planning at 47 mph, and implemented a model-predictive path follower achieving 67 ms latency (IPOPT, CppAD) in C++ for real-time application.

**Princeton University**

NSF RESEARCH ASSISTANT

Princeton, NJ

9/2012-3/2018

Built a rover for Bayesian particle-filter SLAM with LIDAR using Gazebo and ROS, and a custom library for visualizing OpenCV pipelines and dynamic pruning of computational graphs. Simulated thousands of neurons in Numpy and MATLAB, and social agents in OpenMP-accelerated C++. Advised undergrads on theses ranging from power-grid opt. with Gurobi to combinatorial basis extraction.

## SKILLS AND INTERESTS

- Tensorflow/Pytorch
- LangChain and HuggingFace Transformers
- Databricks for data lakes and ML
- Python (15+ yrs), C++, Java
- Plotly, NumPy, SciPy, SymPy, Matplotlib, SKLearn
- Linux, shell scripting, and git
- OPC-UA, MODBUS/serial, and ROS
- Online optimization (IPOPT, CppAD, CasADi)
- Grammatical evolution for code generation
- Bayesian inference and probabilistic computing
- HPC with SLURM and OpenMP, and CUDA
- Differentiable physics for ML
- Mechanistic+data-driven grey-box modeling
- Retrieval-augmented generation (RAG)
- SQL(alchemy), HTML, CSS, JS, Flask, LAMP
- Classical and DL-based computer vision
- Trainee and peer mentoring
- Home automation with Arduino, Raspberry Pi, and 3D printing
- Solo and orchestral violin performance
- Windsurfing and small-boat sailing

## SELECTED PUBLICATIONS AND PRESENTATIONS

**Implementation and (Inverse Modified) Error Analysis for implicitly-templated ODE nets**
A. Zhu, B. Zhu, [Tom Bertalan](#), Y. Tang, I. Kevrekidis

2024

SIAMDS

**Symbolic regression and modular neural diff. eq. for bioproc. engr. and robotics**
[Tom Bertalan](#), J. Lee, Z. Wang, S. Yoon, I. Kevrekidis

2024

MLDS 4

**Certified Invertibility in Neural Networks via Mixed-Integer Programming**
T. Cui, [Tom Bertalan](#), G. Pappas, M. Morari, I. Kevrekidis, M. Fazlyab

2023

L4DC 2023 — PMLR

**Personalized Algorithm Generation: A Case Study in Meta-Learning ODE Integrators**
Y. Guo, F. Dietrich, [Tom Bertalan](#), D. Doncevic, M. Dahmen, I. Kevrekidis, Q. Li

2022

SIAM J. Sci. Comp.

**Local conformal autoencoder for standardized data coordinates**
E. Peterfreund, O. Lindenbaum, F. Dietrich, [Tom Bertalan](#), M. Gavish, I. Kevrekidis, R. Coifman

2020

PNAS

**On Learning Hamiltonian Systems from Data**
[Tom Bertalan](#), F. Dietrich, I. Mezic, and I. Kevrekidis

2019

Chaos

# PUBLICATIONS

---

- Comparing Kinetic versus Stoichiometric Priorities in Hybrid Models of CHO Metabolism** 2026  
P. Khare, N. Ndaheiro, S. Klaubert, E. Ma, [Tom Bertalan](#), Y. Kevrekidis, S. W. Harcum, M. J. Betenbaugh  
*NPJ Sys. Bio. and Appl.*
- Integ. of Bayesian Optim. and Solution Thermo. to Opt. Media Design for Mammalian Biomfg.** 2025  
N. Ndaheiro, E. Ma, [Tom Bertalan](#), M. Donohue, Y. Kevrekidis, and M. Betenbaugh  
*iScience*
- Machine Learning Approaches to Problem Well-Posedness** 2025  
[Tom Bertalan](#), G. Kevrekidis, E. Rebrova, S. Mishra, I. Kevrekidis  
*In preparation*
- Data-driven and Physics Informed Modelling of Chinese Hamster Ovary (CHO) Cell Bioreactors** 2024  
T. Cui, [Tom Bertalan](#), N. Ndaheiro, P. Khare, M. Betenbaugh, C. Maranas, I. Kevrekidis  
*Comp. & Chem. Engr.*
- Implementation and (Inverse Modified) Error Analysis for implicitly-templated ODE nets** 2024  
A. Zhu, B. Zhu, [Tom Bertalan](#), Y. Tang, I. Kevrekidis  
*SIAMDS*
- Transf. establishing equiv. across neural nets.: when have two nets. learned the same task?** 2024  
[Tom Bertalan](#), F. Dietrich, I. Kevrekidis  
*Chaos*
- Certified Invertibility in Neural Networks via Mixed-Integer Programming** 2023  
T. Cui, [Tom Bertalan](#), G. Pappas, M. Morari, I. Kevrekidis, M. Fazlyab  
*L4DC 2023 — PMLR*
- Some of the variables, some of the parameters, some of the times, with some things known: Identification with partial information** 2023  
S. Malani, [Tom Bertalan](#), T. Cui, M. Betenbaugh, J. Avalos, I. Kevrekidis  
*Comp. & Chem. Engr.*
- Learning effective SDEs from microscopic simulations: linking stoch. num. to deep learning** 2023  
F. Dietrich, A. Makeev, G. Kevrekidis, N. Evangelou, [Tom Bertalan](#), S. Reich, I. Kevrekidis  
*Chaos*
- Learning emergent PDEs in a learned emergent space** 2022  
F. Kemeth, [Tom Bertalan](#), T. Thiem, S. Moon, C. Laing, I. Kevrekidis  
*Nature Comm.*
- Personalized Algorithm Generation: A Case Study in Meta-Learning ODE Integrators** 2022  
Y. Guo, F. Dietrich, [Tom Bertalan](#), D. Doncevic, M. Dahmen, I. Kevrekidis, Q. Li  
*SIAM J. Sci. Comp.*
- Initializing LSTM internal states via manifold learning** 2021  
F. Kemeth, [Tom Bertalan](#), N. Evangelou, T. Cui, S. Malani, I. Kevrekidis  
*Chaos*
- Dev. of closures for coarse-scale modeling of multiphase and free surface flows using ML** 2021  
C. Linares, [Tom Bertalan](#), E. Koronaki, Jicai Lu, G. Tryggvason, I. Kevrekidis  
*APS Bulletin*
- Global and local reduced models for interacting, heterogeneous agents** 2021  
T. Thiem, F. Kemeth, [Tom Bertalan](#), C. Liang, I. Kevrekidis  
*Chaos*
- Local conformal autoencoder for standardized data coordinates** 2020  
E. Peterfreund, O. Lindenbaum, F. Dietrich, [Tom Bertalan](#), M. Gavish, I. Kevrekidis, R. Coifman  
*PNAS*
- Emergent spaces for coupled oscillators** 2020  
T. Thiem, M. Kooshkbaghi, [Tom Bertalan](#), C. Liang, I. Kevrekidis  
*Front. in Comp. Neuro.*
- On Learning Hamiltonian Systems from Data** 2019  
[Tom Bertalan](#), F. Dietrich, I. Mezic, and I. Kevrekidis  
*Chaos*
- An Emergent Space for Distributed Data with Hidden Internal Order through Manifold Learning** 2017  
Kemeth, Haugland, Dietrich, [Tom Bertalan](#), Höhle, Li, Bollt, Talmon, Krischer, & Kevrekidis  
*IEEE Access*
- Coarse-grained desc. of dynamics for nets with both intrinsic and structural heterogeneities** 2017  
[Tom Bertalan](#), Y. Wu, C. Laing, C. W. Gear, and I. Kevrekidis.  
*Front. in Comp. Neuro.*
- Dimension reduction in heterogeneous neural networks: Generalized Polynomial Chaos (gPC) and ANalysis-Of-Variance (ANOVA)** 2016  
M. Choi, [Tom Bertalan](#), C. Laing, and I. Kevrekidis.  
*Euro. Phys. J., Special Topics*
- OpenMG: a new multigrid implementation in Python** 2014  
[Tom Bertalan](#), A. Islam, R. Sidje, and E. Carlson  
*Num. Lin. Alg. with App.*

# PRESENTATIONS

---

<b>Integrated Modeling and Optim. of Secondary Drying for Pharma. Freeze-Drying Proc.</b> Tom Bertalan, M. Maritato, A. Ganguly, W. Al Bakri, R. Tukra, B. Lacetti	2025 <i>AIChE 2025 (poster)</i>
<b>Integrated Continuous USP Platform for Max. Productivity and Closed-Loop Controlled CQA</b> J. Lee, Tom Bertalan, Z. Wang, Y. Choi, S. Song, H. Yamanaka, A. Rajendran, J. Han, S. Yoon	2025 <i>AIChE 2025</i>
<b>Digital Twin-Enabled Control of Productivity and CQA in CHO Cell Cultures</b> Y. Choi, S. Yoon, Z. Wang, Tom Bertalan, J. Lee	2025 <i>AIChE 2025 (poster)</i>
<b>Beyond Trial-and-Error: An in silico Model for Accelerated Peristaltic Fill Recipe Development</b> Maritato, Ingram, Sao, Heymann, Bertalan, Banerjee, İçten-Gençer, Ganguly, Pearson, Schlegel	2025 <i>AIChE 2025 (poster)</i>
<b>Symbolic regression and modular neural diff. eq. for bioproc. engr. and robotics</b> Tom Bertalan, J. Lee, Z. Wang, S. Yoon, I. Kevrekidis	2024 <i>MLDS 4</i>
<b>Certified Invertibility in Neural Networks via Mixed-Integer Programming</b> T. Cui, Tom Bertalan, G. Pappas, M. Morari, I. Kevrekidis, M. Fazlyab	2023 <i>Learning for Dyn. Sys.</i>
<b>Coarse-grained and emergent distributed-parameter systems from data</b> H. Arbabi, F. Kemeth, Tom Bertalan, I. Kevrekidis	2021 <i>American Control Conf.</i>
<b>Data-driven model reduction and discovery</b> T. Thiem, Tom Bertalan, F. Kemeth, Y. Psarellis, I. Kevrekidis	2020 <i>AIChE</i>
<b>Dynamical-systems-guided learning of PDEs from data</b> H. Arbabi, Tom Bertalan, A. Roberts, I. Kevrekidis	2020 <i>AIChE</i>
<b>On the data-driven discovery and calibration of closures</b> S. Lee, Y. Psarellis, C. Siettos, Tom Bertalan, D. Amchin, T. Bhattacharjee, S. Datta, I. Kevrekidis	2020 <i>AIChE</i>
<b>Connections between residual networks and explicit numerical integrators, and applications to identification of noninvertible dynamical systems</b> T. Cui, Tom Bertalan, Y. Psarellis, I. Kevrekidis	2020 <i>AIChE</i>
<b>Neural network approach to reduced order modeling of multiphase flows</b> C. Linares, Tom Bertalan, S. Lee, J. Lu, G. Tryggvason, I. Kevrekidis	2020 <i>APS Div. of Fluid Dyn.</i>
<b>PDE+PINN: Learning and Solving a PDE at the Same Time</b> Tom Bertalan, F. Kemeth, T. Cui, I. Kevrekidis	2020 <i>AIChE</i>
<b>Learning Partial Differential Equations from Discrete Space Time Data: Convolutional and Recurrent Networks, and Their Relations to Traditional Numerical Methods</b> Tom Bertalan, F. Dietrich, T. Thiem, R. Farber, I. Kevrekidis, A. Roberts	2019 <i>AIChE</i>
<b>Recurrent Neural Networks, Numerical Integrators and Nonlinear System Identification</b> Tom Bertalan, R. Farber, T. Thiem, F. Dietrich, I. Kevrekidis	2018 <i>AIChE</i>
<b>Coarse-Scale PDEs from Microscopic Observations Via Machine Learning</b> S. Lee, M. Kooshkbaghi, C. Siettos, Tom Bertalan, and I. Kevrekidis	2019 <i>AIChE</i>
<b>When Have Two Nets. Learned the Same Task? Data-Driven Transf. between System Reps.</b> Tom Bertalan, F. Dietrich, T. Thiem, I. Kevrekidis	2019 <i>AIChE</i>
<b>Coarse modeling of circadian rhythms in heterogeneous neural networks</b> Tom Bertalan, C. W. Gear, Y. G. Kevrekidis, M. Henson, E. Herzog, and C. Laing	2017; 2016 <i>Dyn. Days 2017; AIChE</i>
<b>Coarse-graining of heterogeneous neural dynamics</b> Tom Bertalan, M. Choi, C. Laing, I. Kevrekidis	2015 <i>AIChE</i>
<b>Heterogeneity and reduction for complex network dynamics</b> I. Kevrekidis, A. Holiday, Tom Bertalan, and C. Laing	2014 <i>AIChE</i>
<b>Coarse-graining Network Dynamics</b> A. Holiday and Tom Bertalan	2013 <i>Network Front.</i>
<b>nSpyres: an open-source, Python-based framework for sim. of flow through porous media</b> E. Carlson, A. Islam, F. Dumkwu, and Tom Bertalan	2012 <i>Interpore 2012</i>
<b>OpenMG: a new multigrid implementation in Python</b> Tom Bertalan, A. Islam, R. Sidje, and E. Carlson	2012 <i>Proc. 11<sup>th</sup> Python in Sci. Conf.</i>
<b>ESIM: a framework for simulation of dominance hierarchy formation in small animal groups</b> Tom Bertalan and R. Earley	2012 <i>UA Hon. Undergr. Res. Conf.</i>
<b>An open-source computing cluster for virtual experiments with variable parameters</b> Tom Bertalan and E. Carlson	2011 <i>UA Hon. Undergr. Res. Conf.</i>