CAUSAL FIELD THEORY: MECHANISTIC INTERPRETABILITY FOR SPATIO-TEMPORAL BIOLOGICAL SYSTEMS

Arash Mehrjou
Max Planck Institute for Intelligent Systems & GSK.ai
arash@distantvantagepoint.com

Bernhard Schölkopf
Max Planck Institute for Intelligent Systems
bs@tuebingen.mpg.de

ABSTRACT

Understanding how biological interventions propagate through tissue and affect cellular organization is fundamental to drug discovery and therapeutic design. We present Causal Field Theory (CFT), a mathematical framework that extends causal inference to continuous biological fields by modeling causal influence as a propagating field. The framework introduces the response kernel, which quantifies the causal effect of localized interventions such as drug delivery and genetic perturbations, and the causal flux, which measures integrated influence between tissue regions. We demonstrate that tissue-scale interventions can generate greater causal flux than the sum of equivalent cell-level perturbations when aligned with coherent propagation modes. Through numerical experiments on reaction-diffusion systems modeling biological processes, we illustrate how CFT captures causal structure in continuous fields, enabling mechanistic interpretability and principled analysis of interventions in biological systems. This framework addresses a critical gap in current approaches by providing explicit causal semantics for spatio-temporal biological data, enabling lab-in-the-loop systems that reason about intervention effects and guide experimental design.

1 INTRODUCTION

Despite rapid advances in data-driven biology, our limited understanding of the biological mechanisms underlying diseases continues to hinder therapeutic innovation. While genomics and multi-omics platforms have generated vast datasets, translating these into actionable biological insights remains an open challenge (Liu et al., 2024). A fundamental difficulty is that biological interventions, such as drug delivery, genetic perturbations, or immune activation, are typically localized in space, yet their effects propagate through tissue via various mechanisms including diffusion, cell-cell signaling, mechanical coupling, and other transport processes (Jain, 2019; Binnewies et al., 2018; Rosenberger et al., 2023). The resulting causal effects are distributed across space and time, often manifesting as delayed, spatially heterogeneous responses that cannot be captured by classical causal frameworks.

In biological systems, understanding causality requires accounting for how interventions propagate through continuous spatial domainsv. For example, when a drug is applied to a tissue region, the intervention modifies local reaction rates or signaling dynamics, and its effects propagate through the tissue via diffusion, cell-cell communication, and mechanical forces (Orr et al., 2006; Singh et al., 2026). The causal effect at a distant location depends not only on the intervention itself but on the spatial geometry, transport mechanisms, and temporal dynamics of the system. Similar phenomena arise in developmental biology, where morphogen gradients and mechanical forces jointly determine tissue patterning (Grieneisen et al., 2012), and in disease progression, where signals propagate through tissue to affect distant regions (Seydel, 2025).

Existing causal frameworks face fundamental limitations when applied to biological systems (Chevalley et al., 2023; 2025). Potential outcomes approaches assume no interference between units (SUTVA), which fails when spatial coupling enables causal propagation (Rubin, 1974; Imbens & Rubin, 2015). While graphical models such as structural causal models provide a principled frame-

work for causal reasoning, they treat variables as discretely indexed and separable, lacking the geometric structure needed to represent spatial coupling in tissue (Pearl, 2009). Recent work on causal discovery in dynamical systems (Peters et al., 2022; Schreiber, 2000) addresses temporal structure but does not account for spatial propagation. Furthermore, mechanistic simulators reproduce system dynamics with remarkable fidelity but provide no causal semantics; they show *what* happens without explaining *why* or enabling counterfactual reasoning about alternative interventions (Peters et al., 2017; Cartwright et al., 2020). These limitations highlight the need for a framework that can capture both the geometric structure of spatial propagation and the causal semantics required for mechanistic interpretability and counterfactual reasoning in biological systems.

We propose **Causal Field Theory (CFT)**, a framework that treats causality as a property of continuous biological fields rather than finitely many discretely indexed variables. CFT provides causal semantics for biological systems where interventions modify local mechanisms and their effects propagate through space and time. The framework introduces: (1) regional mechanism interventions that modify evolution laws within spatial regions (e.g., localized drug effects, genetic perturbations); (2) causal response kernels that quantify how influence propagates; and (3) derived quantities (causal flux, causal cones) that enable quantitative comparison of interventions and identification of causal pathways.

This paper makes three contributions: (1) a causal formalism for continuous spatio-temporal biological systems that generalizes classical causal inference to field-based dynamics; (2) a theoretical result showing that tissue-scale interventions aligned with coherent propagation modes generate greater causal flux than the sum of equivalent cell-level perturbations; and (3) computational demonstrations validating the framework on reaction-diffusion systems modeling biological processes. CFT addresses a critical gap for lab-in-the-loop systems by providing explicit causal structure for spatio-temporal biological data, enabling mechanistic interpretability, counterfactual reasoning, and principled intervention design.

Figure 1 provides a conceptual overview of CFT, illustrating how it bridges the gap between classical causal frameworks and spatial biological dynamics. The framework generalizes classical causality by introducing regional mechanism interventions and causal response fields that quantify how influence propagates through continuous tissue domains. CFT also generalizes mechanistic modeling approaches such as PDEs, which describe system dynamics but lack causal semantics for interventions. The framework extends mechanistic modeling by providing explicit intervention semantics through regional mechanism edits $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ and quantifying causal effects through response kernels, enabling counterfactual reasoning and intervention design that goes beyond simulation.

2 Causal Field Theory Framework

2.1 System Model and Biological Scope

Consider a biological tissue domain $\mathcal{D} \subset \mathbb{R}^d$ (typically $d = 2$ or $d = 3$) with boundary $\partial\mathcal{D}$. The system state at location $x \in \mathcal{D}$ and time $t \geq 0$ is represented by a field $X(x, t)$, which may be scalar-valued (e.g., morphogen concentration, cytokine level) or vector-valued (e.g., concentrations of multiple species, mechanical displacements). The field evolves according to a local mechanism operator $\mathcal{F}$:

$$\partial_t X(x, t) = \mathcal{F}[X](x, t) + \eta(x, t), \tag{1}$$

where $\mathcal{F}$ encodes intrinsic biological dynamics including reaction-diffusion, cell-cell signaling, and mechanical forces, and $\eta(x, t)$ represents stochastic fluctuations.

In biological systems, $\mathcal{F}$ encodes the intrinsic dynamics governing system evolution, including intracellular processes, intercellular interactions, and environmental influences. For example, in a reaction-diffusion system modeling morphogen gradients, $\mathcal{F}[X](x, t) = D\nabla^2 X(x, t) + f(X(x, t))$ combines spatial coupling through diffusion with local processes such as production and degradation.

From a mathematical perspective, CFT applies to systems where the state field $X(x, t)$ evolves according to a mechanism operator $\mathcal{F}$ that encodes both local dynamics and spatial coupling. The framework requires that $\mathcal{F}$ is sufficiently smooth to ensure the causal response kernel is well-defined, and that the system exhibits spatial coupling enabling causal influence to propagate through space.

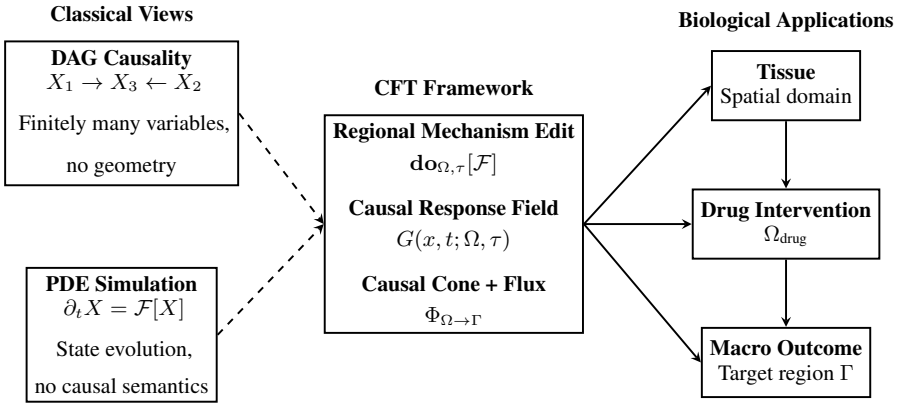


Figure 1: **Conceptual Overview of Causal Field Theory (CFT)**. Left: Classical frameworks: DAG causality (finitely many discretely indexed variables, no geometry) and PDE simulation (state evolution, no causal semantics). Middle: CFT introduces regional mechanism edits $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$, causal response fields $G(x,t;\Omega,\tau)$ that quantify influence propagation, and causal quantities (cones, flux). Right: Biological applications where localized interventions such as drug delivery and genetic perturbations propagate through continuous tissue domains to produce distant effects. CFT bridges the gap between classical causality and spatial biological dynamics, providing causal semantics for continuous spatio-temporal systems.

Biological systems satisfy these properties even though they are intrinsically discrete at the cellular level. When the spacing between cells is small compared to the characteristic length scales of coupling and propagation, systems can be treated as continuous fields. For example, in tissue where cell density is high and signaling molecules diffuse over distances much larger than intercellular spacing, concentration fields and mechanical properties vary smoothly in space, enabling the continuum approximation. Similarly, organoids and cell cultures exhibit continuous field behavior when observed at scales where individual cell boundaries become negligible relative to the spatial extent of causal propagation.

2.2 Regional Mechanism Interventions

Classical causal inference uses interventions that set variables to specific values, such as $\mathbf{do}(X_j = x)$. In biological systems, interventions often modify *mechanisms* rather than variable values. A drug alters reaction rates locally, a gene knockout removes terms from $\mathcal{F}$ in a specific region, or boundary conditions change such as altered nutrient supply or mechanical stretch.

We formalize a *regional mechanism intervention* as an operation $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ that modifies the mechanism operator within a spatial region $\Omega \subset \mathcal{D}$ during a time window around τ. For a finite-duration intervention, the modification occurs for $t \in [\tau, \tau + \Delta]$ where $\Delta > 0$ is the intervention duration. For an instantaneous intervention, we consider the limit $\Delta \to 0$:

$$\mathbf{do}_{\Omega,\tau}[\mathcal{F}]: \quad \mathcal{F}(x,\cdot) \mapsto \mathcal{F}'(x,\cdot) \quad \text{for } x \in \Omega, \quad t \in [\tau, \tau+\Delta], \tag{2}$$

where $\mathcal{F}'$ is the modified mechanism operator. Outside Ω and outside the time window, $\mathcal{F}$ remains unchanged. This generalizes the do-operator from variables to spatially localized mechanisms, embedding interventions directly into the system's dynamics. In biological contexts, $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ might represent localized drug application, genetic perturbation in a tissue region, or mechanical modulation of a specific domain.

2.3 Causal Response Kernel

We define the *causal response kernel* $G(x,t;\Omega,\tau)$ as the functional derivative of the field with respect to the intervention:

$$G(x,t;\Omega,\tau) := \frac{\delta X(x,t)}{\delta\, \mathbf{do}_{\Omega,\tau}[\mathcal{F}]}. \tag{3}$$

The causal response kernel quantifies how much an intervention $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ affects the field at location x and time $t > \tau$. Intuitively, G answers the question of how much the intervention mattered at a given location and time. The response kernel is itself a field that, for each (x, t), gives a scalar or vector measuring the sensitivity of $X(x, t)$ to the intervention.

Theorem 1 (Well-definedness of the Causal Response Kernel)**.** *Under mild regularity assumptions on $\mathcal{F}$ (Fréchet-differentiability and smooth dependence of solutions on parameters), the causal response kernel $G(x, t; \Omega, \tau) = \delta X(x, t)/\delta\, \mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ exists, is unique, and satisfies the linearized evolution equation:*

$$\partial_t G(x, t; \Omega, \tau) = D\mathcal{F}[X] \cdot G(x, t; \Omega, \tau) + S_\Omega(x, t), \tag{4}$$

where $D\mathcal{F}[X]$ is the Fréchet derivative of $\mathcal{F}$ evaluated along the baseline trajectory X, and $S_\Omega(x, t)$ is a localized source term induced by the intervention in region Ω.

A proof sketch is provided in Appendix A.4.

This theorem establishes that G is a well-defined mathematical object, connecting CFT to Green's functions (Evans, 2010), linear response theory (Oppenheim et al., 1997; Chen, 1984), and adjoint sensitivity analysis. The response kernel provides a rigorous foundation for quantifying causal effects in biological systems, enabling mechanistic interpretability by revealing how interventions propagate through tissue.

2.4 Derived Causal Quantities

From the response kernel G, we compute several quantities that make causal influence measurable and comparable. We define the *causal flux* from region Ω to region Γ as:

$$\Phi_{\Omega\to\Gamma} = \int_\Gamma \int_\tau^\infty G(x, t; \Omega, \tau)\, dt\, dx. \tag{5}$$

This measures the total integrated influence that an intervention in Ω exerts on Γ over all time, enabling quantitative comparison of causal pathways. For example, we can compare $\Phi_{\Omega_1\to\Gamma}$ and $\Phi_{\Omega_2\to\Gamma}$ to determine which intervention region has stronger causal influence on a target tissue region.

We now define the *causal cone* as the support of G: the set of (x, t) for which $G(x, t; \Omega, \tau) \neq 0$. In biological systems with finite propagation speeds (e.g., diffusion with bounded coefficients, wave equations for action potentials), the causal cone has a well-defined boundary determined by the maximum speed of influence propagation.

Theorem 2 (Finite-Speed Propagation of Causal Influence)**.** *If $\mathcal{F}$ satisfies locality and finite-speed propagation (e.g., parabolic or hyperbolic PDEs with bounded coefficients), then $G(x, t; \Omega, \tau) = 0$ outside a causal cone. Specifically, if the maximum propagation speed is c, then $G(x, t; \Omega, \tau) = 0$ for $|x - x_\Omega| > c(t - \tau)$, where x_Ω is any point in Ω.*

This theorem formally justifies the causal cone concept and distinguishes CFT from DAG-based causality (which has no geometry) and potential outcomes (which assumes no propagation). In biological systems, the causal cone reveals which tissue regions can be causally affected by an intervention and by when, providing crucial information for experimental design and therapeutic intervention planning. A proof sketch is provided in Appendix A.5.

3 Scale Alignment and Causal Effectiveness

In spatially extended biological systems, interventions can be defined at multiple spatial scales, from fine-grained perturbations of individual cells to coarse-grained modifications of tissue regions, organs, or organ systems. A fundamental question arises: *what determines whether coarse-grained (macro) or fine-grained (micro) interventions exert stronger or more robust causal influence on downstream system behavior?*

CFT addresses this through the notion of *causal coherence*, which characterizes regimes where interventions defined at a coarser scale induce responses that are more structured, robust, or spatially

extensive than those obtained by aggregating corresponding fine-scale interventions. The key insight is that causal effectiveness depends on alignment between the scale of intervention and the system's dominant spatio-temporal propagation modes. Micro-level mechanisms generate the dynamics; macro-level effectiveness arises from geometric and dynamical alignment with coherent propagation structures. Detailed discussion is provided in Appendix A.1.

We formalize this through a comparison of macro- and micro-interventions in a linearized spatio-temporal system. Consider a spatial region Ω partitioned into n disjoint subregions $\{\Omega_i\}_{i=1}^n$. We define the causal flux $\Phi_{\Omega\to\Gamma} = \int_\Gamma \int_{t>\tau} |G(x,t;\Omega,\tau)|\, dt\, dx$ as a measure of integrated causal influence from intervention region Ω to target region Γ.

Proposition 1 (Scale Alignment and Causal Flux)**.** *Let Ω be a spatial region partitioned into n disjoint subregions $\{\Omega_i\}_{i=1}^n$. Consider a macro-intervention $\mathbf{do}_{\Omega,\tau}[\mathcal{F}_{\text{macro}}]$ that modifies the mechanism operator uniformly over Ω at time τ, and micro-interventions $\{\mathbf{do}_{\Omega_i,\tau}[\mathcal{F}_{\text{micro},i}]\}_{i=1}^n$ that modify the mechanism operator on each subregion Ω_i at time τ.*

Assume: (1) The system is in the linearized regime around a baseline trajectory $X_0(x,t)$; (2) The macro-intervention satisfies $\mathcal{F}_{\text{macro}}[X_0](x,t) = \sum_{i=1}^n \chi_{\Omega_i}(x)\mathcal{F}_{\text{micro},i}[X_0](x,t)$ for $x \in \Omega$; (3) The response kernels exist and are well-defined; (4) The system exhibits spatial coherence with correlation length ℓ_c such that cross-region contributions to the response kernel are non-negative; (5) The intervention scale satisfies $\text{diam}(\Omega) \gtrsim \ell_c$.

Then, for any target region Γ:

$$\Phi^{\text{macro}}_{\Omega\to\Gamma} \geq \sum_{i=1}^n \Phi^{\text{micro},i}_{\Omega_i\to\Gamma}, \tag{6}$$

where $\Phi^{\text{macro}}_{\Omega\to\Gamma}$ and $\Phi^{\text{micro},i}_{\Omega_i\to\Gamma}$ are the causal fluxes from macro and micro interventions, respectively.

A proof sketch is provided in Appendix A.6.

The proof leverages constructive interference. When micro-interventions are spatially correlated over scales comparable to or larger than the system's correlation length, their effects combine coherently. Macro-interventions that align with these correlated modes exploit this coherence, whereas isolated micro-interventions exhibit destructive interference or decay before reaching the target region.

This result has direct biological relevance for intervention design. In drug delivery, regional administration that matches tissue-scale diffusion and transport properties produces more robust therapeutic effects than highly localized, fragmented applications. For example, drug delivery to tissue regions that align with coherent propagation modes exploits diffusion and cell-cell signaling to generate spatially extensive responses, whereas fragmented local applications may fail to create coherent tissue-scale effects. Similarly, tissue-scale mechanical interventions produce more robust long-range effects on cell behavior and tissue remodeling than highly localized mechanical perturbations (Orr et al., 2006), as mechanical stress propagates through tissue with characteristic length scales that regional interventions can exploit.

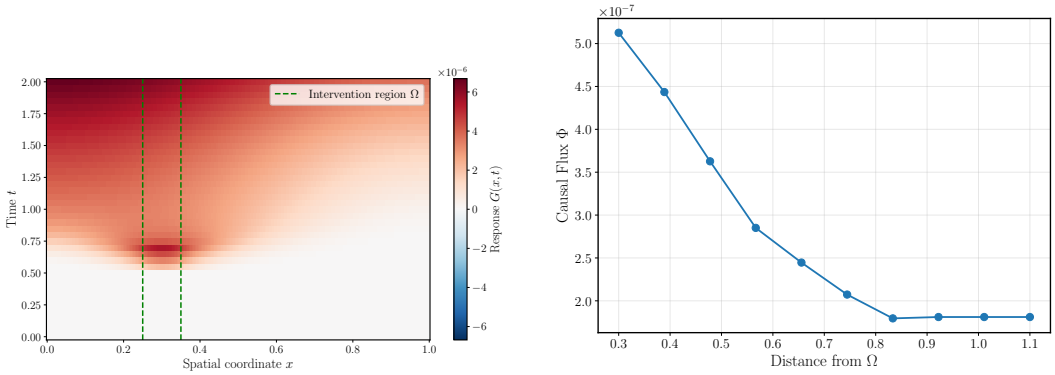
These examples illustrate that interventions aligned with the system's characteristic correlation lengths and propagation modes can generate greater causal flux than fragmented applications. This suggests that the scale of intervention relative to system correlation lengths may be an important consideration in experimental design, therapeutic intervention design, and multiscale modeling. Detailed discussion is provided in Appendix A.1.

4 Computational Demonstrations

We present two computational experiments demonstrating CFT's capabilities on reaction-diffusion systems modeling biological processes. All simulations solve the PDE $\partial_t X = D\nabla^2 X + rX(1 - X/K) - \lambda X$ with appropriate boundary conditions, where $X(x,t)$ represents a morphogen or cytokine-like concentration field. These experiments validate the theoretical framework and illustrate how CFT captures causal structure in continuous biological fields. The reaction-diffusion model serves as a prototypical example of spatio-temporal dynamics found across biological systems, from morphogen gradients in development to cytokine signaling in immune responses.

4.1 Experiment 1: Causal Cone and Flux

Our first experiment demonstrates the fundamental causal objects of CFT in a one-dimensional reaction-diffusion system. We apply a regional mechanism intervention by locally increasing the production rate r within a region Ω for a short time window, then track how the causal effect propagates through space and time.



(a) Causal response kernel heatmap

(b) Causal flux vs distance

Figure 2: **Experiment 1: Causal Response Kernel and Flux**. (a) Heatmap of the response kernel $G(x,t)$ showing spatio-temporal propagation of causal influence from a regional intervention. The causal cone (green dashed lines) delineates the intervention region Ω. (b) Causal flux $\Phi_{\Omega\to\Gamma}$ as a function of distance, demonstrating how causal influence decays as it propagates through space.

Figure 2a shows the response kernel as a heatmap, revealing the causal cone, which is the region of space-time where the intervention has non-negligible effect. The cone expands over time as influence diffuses outward from the intervention region. Figure 2b shows how causal flux decreases with distance, quantifying how effectively causal influence reaches distant target regions. This enables quantitative comparison of intervention strategies for biological systems.

4.2 Experiment 2: Geometry Effects on Causal Propagation

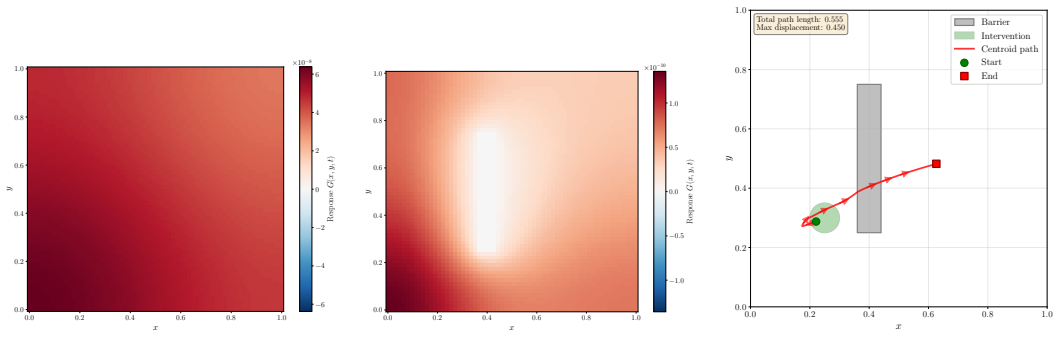
Our second experiment demonstrates how tissue geometry affects causal propagation, specifically examining how barriers or obstacles influence propagation. We compare causal propagation in a homogeneous 2D domain versus a domain containing a barrier that simulates a scar, obstacle, or impermeable region in tissue.

Figure 3 shows how barriers alter causal propagation. The response in the barrier case exhibits deformation and cannot penetrate the obstacle, indicating that causal influence follows the geometry of the system. This demonstrates that causal influence is geometrically mediated and can be blocked, reflected, or channeled by spatial structure, in contrast to graphical models which lack explicit geometric structure. In biological systems, barriers such as scars, blood vessels, or cell layers can shield some regions from causal influence while channeling it to others, which may be relevant for understanding intervention effects in tissue.

5 Discussion and Outlook

5.1 Relation to Existing Causal Frameworks

As illustrated in Figure 1, CFT generalizes and extends classical causal frameworks in several important ways. In the well-mixed limit where spatial coupling becomes infinitely fast, CFT reduces to standard causal effects on aggregated variables, recovering the potential outcomes framework with no interference (SUTVA becomes vacuously true as spatial structure disappears). Under spatial discretization, CFT reduces to dynamic structural causal models with explicitly defined interference structure. In the linearized regime, the response kernel coincides with Green's functions from linear systems theory, connecting CFT to classical sensitivity analysis. These reductions demonstrate



(a) Homogeneous (b) With barrier (c) Centroid trajectory

Figure 3: **Experiment 2: Geometry Effects**. (a) Response kernel in homogeneous domain showing radial propagation. (b) Response kernel with barrier (dark region) showing deformation around obstacle. (c) Centroid trajectory of the causal flux mass; notice that the centroid can pass through the obstacle because it is the center of mass of the distributed influence (which propagates around the barrier), not a physical particle trajectory.

that CFT is a strict generalization rather than a competing alternative, providing causal semantics precisely in regimes where existing approaches become insufficient.

The key distinction is that CFT treats causality as a geometric, propagating property rather than an instantaneous relationship between discretely indexed variables. This perspective is essential for biological systems where spatial coupling enables causal influence to propagate through continuous domains, creating distributed, history-dependent effects that cannot be captured by DAGs or potential outcomes frameworks. The causal cone concept formalizes this geometric structure, providing a principled way to identify which regions can be causally affected by an intervention and by when. This geometric perspective enables CFT to capture phenomena such as wave propagation, diffusion-mediated signaling, and mechanical stress transfer, where causal influence has explicit spatial structure that determines how interventions propagate through the system. The framework's ability to represent geometric constraints on causal propagation, such as barriers that block or redirect influence, distinguishes it from graph-based approaches that lack explicit spatial structure.

CFT provides a mathematical foundation for causal reasoning in continuous spatio-temporal biological systems, reframing causality as a dynamic, geometric property where influence propagates through space and time according to system dynamics. The framework is particularly timely given recent technological advances. Spatial omics technologies provide high-resolution snapshots of biological systems (Liu et al., 2024; Rosenberger et al., 2023; Mehrizi et al., 2023), perturbation screens enable systematic exploration of localized interventions (Jinek et al., 2012; Miladinovic et al., 2025; **?**), and agent-based simulators generate synthetic spatio-temporal data (Ghaffarizadeh et al., 2018).

5.2 Applications in Lab-in-the-Loop Systems

CFT provides essential causal semantics for lab-in-the-loop systems that reason about biological interventions and guide experimental design (Mehrjou et al., 2021; Lyle et al., 2023). In particular, the causal response kernel $G(x, t; \Omega, \tau)$ provides mechanistic interpretability by revealing how interventions propagate through tissue and affect distant regions, enabling systems to explain intervention effects in terms of spatial propagation mechanisms. Regional mechanism edits $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ enable counterfactual reasoning for experimental design, allowing systems to answer questions about alternative intervention strategies. By computing causal flux $\Phi_{\Omega\to\Gamma}$ for different intervention strategies, systems can identify which interventions maximize desired effects.

Proposition 1 provides guidance on which scales of representation are most causally effective, suggesting that causal structure should be represented at scales that align with coherent propagation modes rather than at the finest available resolution. This connects to recent work on multi-scale architectures in biological machine learning (Li et al., 2020; Lu et al., 2019).

The framework can be integrated with operator learning approaches (Li et al., 2020; Lu et al., 2019) and sparse discovery methods (Brunton et al., 2016; Rudy et al., 2017) that learn interpretable mechanism operators from biological data. Once a mechanism operator $\mathcal{F}$ is learned or specified, CFT provides the mathematical framework to analyze its causal implications. The response kernel also provides a natural framework for quantifying uncertainty in causal effects, enabling probabilistic causal inference that accounts for epistemic and aleatoric uncertainty.

5.3 Data-Driven Implementation

In practice, CFT can be applied to biological systems where the mechanism operator $\mathcal{F}$ is learned from observational or interventional data rather than being fully specified a priori. The framework supports a hybrid approach where different components can be learned or partially learned depending on available data and domain knowledge.

The mechanism operator $\mathcal{F}$ can be learned using neural operators (Li et al., 2020; Lu et al., 2019) that approximate the evolution dynamics from spatio-temporal data. Alternatively, sparse discovery methods (Brunton et al., 2016; Rudy et al., 2017) can identify interpretable PDE forms for $\mathcal{F}$ from data. Once $\mathcal{F}$ is learned, the causal response kernel $G(x, t; \Omega, \tau)$ can be computed by solving the linearized evolution equation (Theorem 1), which requires computing the Fréchet derivative $D\mathcal{F}[X_0]$ along the baseline trajectory.

For systems where the full state field $X(x, t)$ is not directly observable, latent variable models can be used to infer the state from partial observations. The learned mechanism operator can then be used to compute causal quantities such as the response kernel and causal flux. When measurement noise or parameter uncertainty is present, stochastic extensions enable uncertainty quantification by propagating uncertainty through the response kernel computation.

The causal flux $\Phi_{\Omega \to \Gamma}$ can be computed from the response kernel and used to compare intervention strategies. For intervention design, the framework enables optimization over intervention regions Ω and times τ to maximize causal flux to target regions Γ while respecting constraints. Algorithm 1 provides a practical workflow for data-driven causal inference using CFT.

References

Mikhail Binnewies, Edward W Roberts, Kelly Kersten, Vincent Chan, Douglas F Fearon, Miriam Merad, Lisa M Coussens, Dmitry I Gabrilovich, Suzanne Ostrand-Rosenberg, Catherine C Hedrick, et al. Understanding the tumor immune microenvironment (time) for effective therapy. *Nature medicine*, 24(5):541–550, 2018.

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

Nancy Cartwright, John Pemberton, and Sarah Wieten. Mechanisms, laws and explanation. *European Journal for Philosophy of Science*, 10(3):25, 2020.

Chi-Tsong Chen. *Linear system theory and design*. Saunders college publishing, 1984.

Mathieu Chevalley, Jacob Sackett-Sanders, Yusuf Roohani, Pascal Notin, Artemy Bakulin, Dariusz Brzezinski, Kaiwen Deng, Yuanfang Guan, Justin Hong, Michael Ibrahim, et al. The causalbench challenge: A machine learning contest for gene network inference from single-cell perturbation data. *arXiv preprint arXiv:2308.15395*, 2023.

Mathieu Chevalley, Yusuf H Roohani, Arash Mehrjou, Jure Leskovec, and Patrick Schwab. A large-scale benchmark for network inference from single-cell perturbation data. *Communications Biology*, 8(1):412, 2025.

Lawrence C Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2010.

Ahmadreza Ghaffarizadeh, Randy Heiland, Samuel H Friedman, Shannon M Mumenthaler, and Paul Macklin. Physicell: an open source physics-based cell simulator for 3-d multicellular systems. *PLoS computational biology*, 14(2):e1005991, 2018.

Verônica A Grieneisen, Ben Scheres, Paulien Hogeweg, and Athanasius F M Marée. Morphogengineering roots: comparing mechanisms of morphogen gradient formation. *BMC systems biology*, 6(1):37, 2012.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

Kewal K Jain. An overview of drug delivery systems. *Drug delivery systems*, pp. 1–54, 2019.

Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A programmable dual-rna–guided dna endonuclease in adaptive bacterial immunity. *science*, 337(6096):816–821, 2012.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Longqi Liu, Ao Chen, Yuxiang Li, Jan Mulder, Holger Heyn, and Xun Xu. Spatiotemporal omics for biology and medicine. *Cell*, 187(17):4488–4519, 2024.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

Clare Lyle, Arash Mehrjou, Pascal Notin, Andrew Jesson, Stefan Bauer, Yarin Gal, and Patrick Schwab. Discobax: Discovery of optimal intervention sets in genomic experiment design. In *International Conference on Machine Learning*, pp. 23170–23189. PMLR, 2023.

Rahil Mehrizi, Arash Mehrjou, Maryana Alegro, Yi Zhao, Benedetta Carbone, Carl Fishwick, Johanna Vappiani, Jing Bi, Siobhan Sanford, Hakan Keles, et al. Multi-omics prediction from high-content cellular imaging with deep learning. *arXiv preprint arXiv:2306.09391*, 2023.

Arash Mehrjou, Ashkan Soleymani, Andrew Jesson, Pascal Notin, Yarin Gal, Stefan Bauer, and Patrick Schwab. Genedisco: A benchmark for experimental design in drug discovery. 2021.

Djordje Miladinovic, Tobias Höppe, Mathieu Chevalley, Andreas Georgiou, Lachlan Stuart, Arash Mehrjou, Marcus Bantscheff, Bernhard Schölkopf, and Patrick Schwab. In silico biological discovery with large perturbation models. *Nature Computational Science*, pp. 1–12, 2025.

Alan V Oppenheim, Alan S Willsky, and Syed Hamid Nawab. *Signals & systems*. Pearson Educación, 1997.

A Wayne Orr, Brian P Helmke, Brett R Blackman, and Martin A Schwartz. Mechanisms of mechanotransduction. *Developmental cell*, 10(1):11–20, 2006.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT press, 2017.

Jonas Peters, Stefan Bauer, and Niklas Pfister. Causal models for dynamical systems. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pp. 671–690. 2022.

Florian A Rosenberger, Marvin Thielert, Maximilian T Strauss, Lisa Schweizer, Constantin Ammar, Sophia C Mädler, Andreas Metousis, Patricia Skowronek, Maria Wahle, Katherine Madden, et al. Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *Nature Methods*, 20(10):1530–1536, 2023.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science advances*, 3(4):e1602614, 2017.

Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

Caroline Seydel. Beyond cell atlases: spatial biology reveals mechanisms behind disease. *Nature biotechnology*, 43(6):841–844, 2025.

Amogh Singh, Patrick Schwab, Kim Branson, Pietro Lio, and Arash Mehrjou. Nexus: A multi-scale simulator for biological control and causal discovery. In *ICLR 2026 Workshop on Machine Learning for Genomics Explorations*, 2026.

A Appendix

A.1 Scale Alignment and Causal Effectiveness: Detailed Discussion

A.1.1 Problem Framing

In spatially extended dynamical systems, interventions can be defined at multiple spatial scales, from fine-grained local perturbations to coarse-grained regional modifications. Intuition might suggest that finer resolution provides greater control, yet this is not necessarily the case. In systems with spatial coupling, the effectiveness of an intervention depends not only on its magnitude but on how well it aligns with the system's dominant spatio-temporal propagation modes. Finer-grained interventions may fail to exploit coherent propagation structures, while coarser interventions that match these structures can produce more robust and spatially extensive effects.

This question is particularly relevant in biological systems, where interventions may target individual cells, tissue regions, organs, or organ systems. Understanding which scale of intervention maximizes causal effectiveness is essential for experimental design, therapeutic development, and multiscale modeling.

A.1.2 Conceptual Framework

This framework explicitly distinguishes scale alignment from philosophical notions of emergence or downward causation. We do not claim that macro-level causes operate independently of micro-level mechanisms. Rather, we show that when micro-level mechanisms generate coherent propagation modes at larger scales, interventions aligned with these modes are more causally effective. The micro-level mechanisms still generate the dynamics; macro-level effectiveness arises from geometric and dynamical alignment. The spatial coherence assumption captures the idea that the system's propagation modes have characteristic length scales, and interventions aligned with these scales exploit coherent propagation.

A.1.3 Interpretation of the Result

This result makes conditional claims under explicit assumptions. It does not assert that macro-interventions are always more effective; rather, it provides sufficient conditions under which coordinated coarse-grained interventions produce greater integrated causal influence than uncoordinated fine-scale interventions.

The result does not claim that micro-level mechanisms are irrelevant or that macro-level causes operate independently. Micro-level mechanisms still generate the dynamics; macro-level effectiveness arises from geometric and dynamical alignment. The result is mathematically rigorous but modest: it applies to linearized systems with well-defined response kernels and requires explicit coherence conditions. These conditions are satisfied in many biological systems where spatial coupling creates coherent propagation modes, but the result does not claim universal applicability.

A.1.4 Biological Examples

In developmental biology, morphogen gradients exhibit coherent propagation modes with characteristic length scales determined by diffusion coefficients and degradation rates. Tissue-scale interventions that align with these modes exploit coherent propagation, while cell-level perturbations may fail to generate coherent responses.

Tissue-scale mechanical or stiffness interventions produce more robust and predictable long-range effects on cell behavior, gene expression, and tissue remodeling than cell-autonomous mechanical

perturbations. Mechanical stress propagates through tissue with characteristic length scales determined by material properties and cell-cell adhesion. Regional interventions that match these scales generate coherent mechanical fields, while isolated cell-level perturbations decay before reaching distant targets.

Inflammatory or signaling fields in tissues provide another example. Tissue-scale inflammatory fronts can drive system-level outcomes more predictably than the release of individual cytokines. Cytokine signaling exhibits coherent propagation through diffusion and cell-cell communication, with correlation lengths determined by diffusion coefficients and receptor binding kinetics. Regional interventions that align with these scales exploit coherent propagation, while point sources may fail to generate system-wide responses.

These examples instantiate the assumptions of the mathematical result: they involve diffusion and spatial coupling that create correlation lengths, and interventions at scales comparable to or larger than these correlation lengths exploit coherent propagation modes.

A.1.5 Practical Implications

This framework has direct implications for experimental design. When choosing intervention scale, researchers should consider the system's characteristic correlation lengths and propagation modes. Interventions aligned with these scales are more likely to produce robust, predictable, and spatially extensive effects. For example, in studies of tissue morphogenesis, tissue-scale mechanical interventions may be more effective than cell-level perturbations for producing long-range effects.

For drug and therapeutic intervention design, the framework suggests that regional targeting may be more effective than cell-autonomous targeting when the therapeutic mechanism exploits coherent propagation modes. This is particularly relevant for interventions that modify signaling pathways, mechanical properties, or metabolic processes with characteristic spatial scales. Regional drug delivery that matches these scales may produce more robust therapeutic effects than cell-level targeting.

For multiscale modeling and abstraction choices, the framework suggests that causal structure should be represented at scales that align with coherent propagation modes rather than at the finest available resolution. This provides a principled basis for choosing appropriate levels of description in multiscale models, balancing computational efficiency with causal accuracy.

The emphasis on robustness, predictability, and efficiency of interventions makes this framework particularly valuable for lab-in-the-loop systems, where interventions must be designed to produce reliable and interpretable effects.

A.2 Data-Driven Causal Inference Algorithm

Algorithm 1 outlines a practical workflow for applying CFT to biological systems with learned mechanism operators.

The algorithm requires solving PDEs in Steps 2 and 4, which can be done using standard numerical methods such as finite difference, finite element, or spectral methods. For high-dimensional systems, neural operators can also be used to approximate the solution operators. The computational complexity depends on the spatial discretization and temporal resolution, but the framework scales to systems with thousands of spatial points and reasonable time horizons.

A.3 Proof Sketches

A.4 Proof of Theorem 1: Well-definedness of the Causal Response Kernel

We provide a sketch of the proof that the causal response kernel $G(x, t; \Omega, \tau) = \delta X(x, t)/\delta\, \mathbf{do}_{\Omega,\tau}[\mathcal{F}]$ is well-defined under mild regularity assumptions on $\mathcal{F}$.

Consider the baseline system evolution:

$$\partial_t X(x, t) = \mathcal{F}[X](x, t), \tag{7}$$

with initial condition $X(x, 0) = X_0(x)$.

Algorithm 1 Data-Driven Causal Inference with CFT

Require: Observational or interventional data $\mathcal{D} = \{(X_i(x,t), \Omega_i, \tau_i)\}_{i=1}^n$
Ensure: Causal response kernels $G(x,t;\Omega,\tau)$ and causal flux $\Phi_{\Omega\to\Gamma}$
Step 1: Learn mechanism operator
Learn $\mathcal{F}$ from data using neural operators or sparse discovery
Validate learned operator on held-out data
Step 2: Compute baseline trajectory
Solve $\partial_t X_0(x,t) = \mathcal{F}[X_0](x,t)$ with initial conditions
Store baseline trajectory $X_0(x,t)$ for all (x,t)
Step 3: Compute Fréchet derivative
Evaluate $D\mathcal{F}[X_0]$ along baseline trajectory
Construct linearized evolution operator
Step 4: Compute response kernel
for each intervention (Ω,τ) **do**
    Solve $\partial_t G = D\mathcal{F}[X_0]\cdot G + S_\Omega$ with initial conditions
    Store $G(x,t;\Omega,\tau)$ for all (x,t)
end for
Step 5: Compute causal quantities
for each target region Γ **do**
    Compute causal flux $\Phi_{\Omega\to\Gamma} = \int_\Gamma \int_{t>\tau} |G(x,t;\Omega,\tau)|\, dt\, dx$
    Identify causal cone boundaries
end for
Step 6: Intervention design (optional)
Optimize over (Ω,τ) to maximize $\Phi_{\Omega\to\Gamma}$ subject to constraints
Return optimal intervention strategy

Under a regional mechanism intervention $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$, the mechanism operator is modified to $\mathcal{F}'$ in region Ω:

$$\mathcal{F}'(x,\cdot) = \begin{cases} \mathcal{F}(x,\cdot) + \delta\mathcal{F}(x,\cdot) & \text{for } x \in \Omega, \\ \mathcal{F}(x,\cdot) & \text{otherwise,} \end{cases} \tag{8}$$

where $\delta\mathcal{F}$ represents the mechanism perturbation.

Let $X_\epsilon(x,t)$ denote the solution under the perturbed mechanism, parameterized by intervention magnitude ϵ. The causal response kernel is defined as:

$$G(x,t;\Omega,\tau) = \lim_{\epsilon\to 0} \frac{X_\epsilon(x,t) - X_0(x,t)}{\epsilon} = \frac{\delta X(x,t)}{\delta\, \mathbf{do}_{\Omega,\tau}[\mathcal{F}]}. \tag{9}$$

Under the assumption that $\mathcal{F}$ is Fréchet-differentiable, the perturbation $\delta X = X_\epsilon - X_0$ satisfies a linearized variational equation. Taking the limit $\epsilon \to 0$, we obtain:

$$\partial_t G(x,t;\Omega,\tau) = D\mathcal{F}[X_0]\cdot G(x,t;\Omega,\tau) + S_\Omega(x,t), \tag{10}$$

where $D\mathcal{F}[X_0]$ is the Fréchet derivative of $\mathcal{F}$ evaluated along the baseline trajectory X_0, and $S_\Omega(x,t)$ is a localized source term:

$$S_\Omega(x,t) = \begin{cases} \frac{\delta\mathcal{F}[X_0](x,t)}{\delta\epsilon}\big|_{\epsilon=0} & \text{for } x \in \Omega,\ t \in [\tau-\epsilon, \tau+\epsilon], \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The initial condition for G is $G(x,\tau;\Omega,\tau) = 0$ (no response before the intervention), and the boundary conditions match those of the original system.

By standard results on linear PDEs and variational equations (Evans, 2010), the solution G exists and is unique provided: (1) The Fréchet derivative $D\mathcal{F}[X_0]$ is well-defined and bounded; (2) The source term S_Ω is locally integrable; (3) The baseline solution X_0 has sufficient regularity (typically C^1 in time and C^2 in space for parabolic systems). These conditions are satisfied under the stated regularity assumptions on $\mathcal{F}$. The solution G inherits the regularity properties of the underlying dynamics, establishing that G is a well-defined mathematical object.

A.5 Proof of Theorem 2: Finite-Speed Propagation of Causal Influence

We prove that if $\mathcal{F}$ satisfies locality and finite-speed propagation, then $G(x, t; \Omega, \tau) = 0$ outside a causal cone. The proof follows from standard PDE propagation bounds: the response kernel inherits the propagation characteristics of the underlying dynamics.

The response kernel G satisfies the linearized evolution equation:

$$\partial_t G(x, t; \Omega, \tau) = D\mathcal{F}[X_0] \cdot G(x, t; \Omega, \tau) + S_\Omega(x, t), \tag{12}$$

where $S_\Omega(x, t)$ is a localized source term with support in $\Omega \times [\tau, \tau + \Delta]$.

For hyperbolic systems with finite propagation speeds, standard results (Evans, 2010) establish that solutions propagate with bounded speed. Specifically, if the maximum propagation speed is c, then the solution G has support contained in the set $\{(x, t) : |x - x_\Omega| \leq c(t - \tau)\}$ for any $x_\Omega \in \Omega$. This follows from the fact that the linearized operator $D\mathcal{F}[X_0]$ encodes the same propagation characteristics as the original mechanism operator $\mathcal{F}$.

For parabolic systems such as reaction-diffusion equations, the propagation speed is effectively infinite due to diffusive coupling. The response kernel is nonzero everywhere for $t > \tau$, but decays rapidly with distance. For example, in the heat equation, the Green's function has Gaussian decay $G(x, t) \sim \exp(-|x - x_\Omega|^2/(4D(t - \tau)))$, where D is the diffusion coefficient. While influence spreads instantaneously, it is strongly attenuated at large distances, effectively creating a soft causal boundary determined by the decay rate rather than a strict support condition.

The causal cone is defined as the set where G has significant magnitude, providing a principled way to identify which regions can be causally affected by an intervention and by when. For hyperbolic systems, this is a strict support condition; for parabolic systems, it is determined by the decay rate of the response kernel. This geometric structure distinguishes CFT from graph-based causal frameworks that lack explicit spatial structure.

A.6 Proof Sketch: Scale Alignment and Causal Flux

We sketch the proof of Theorem 1. By assumption (1), we are in the linearized regime around baseline trajectory $X_0(x, t)$. In this regime, the response kernel can be expressed using the Green's function $K(x, x', t - \tau)$ of the linearized system. The macro-intervention response kernel is:

$$G_{\text{macro}}(x, t; \Omega, \tau) = \int_\Omega K(x, x', t - \tau) \delta\mathcal{F}_{\text{macro}}[X_0](x', \tau)\, dx', \tag{13}$$

where $\delta\mathcal{F}_{\text{macro}}$ represents the perturbation to the mechanism operator. Similarly, for micro-interventions:

$$G_{\text{micro},i}(x, t; \Omega_i, \tau) = \int_{\Omega_i} K(x, x', t - \tau) \delta\mathcal{F}_{\text{micro},i}[X_0](x', \tau)\, dx'. \tag{14}$$

By assumption (2), when the macro-intervention is applied uniformly, $\delta\mathcal{F}_{\text{macro}}[X_0](x, t) = \sum_{i=1}^n \chi_{\Omega_i}(x) \delta\mathcal{F}_{\text{micro},i}[X_0](x, t)$ for $x \in \Omega$. Substituting this into the expression for G_{macro}:

$$\begin{aligned} G_{\text{macro}}(x, t; \Omega, \tau) &= \int_\Omega K(x, x', t - \tau) \sum_{i=1}^n \chi_{\Omega_i}(x') \delta\mathcal{F}_{\text{micro},i}[X_0](x', \tau)\, dx' \\ &= \sum_{i=1}^n \int_{\Omega_i} K(x, x', t - \tau) \delta\mathcal{F}_{\text{micro},i}[X_0](x', \tau)\, dx' \\ &= \sum_{i=1}^n G_{\text{micro},i}(x, t; \Omega_i, \tau). \end{aligned} \tag{15}$$

At the level of response kernels, linearity implies equality. However, the inequality arises when computing causal flux, which integrates G over space and time. By assumption (4), the spatial coherence condition ensures that when interventions are applied simultaneously (as in the macro case), cross-region contributions are non-negative, leading to constructive interference. This coherence is lost when interventions are applied sequentially or independently, yielding the inequality $\Phi^{\text{macro}}_{\Omega \to \Gamma} \geq \sum_{i=1}^n \Phi^{\text{micro},i}_{\Omega_i \to \Gamma}$.

B KEY OBJECTS IN CFT

For reference, we provide a glossary of key objects in CFT:

- **State field** $X(x,t)$: The system state at location $x \in \mathcal{D}$ and time $t \geq 0$, representing concentrations, densities, or other field variables in biological systems.
- **Mechanism operator** $\mathcal{F}$: The functional encoding intrinsic dynamics: $\partial_t X(x,t) = \mathcal{F}[X](x,t)$.
- **Regional intervention** $\mathbf{do}_{\Omega,\tau}[\mathcal{F}]$: An operation that modifies the mechanism operator within spatial region Ω during a time window around τ.
- **Causal response kernel** $G(x,t;\Omega,\tau)$: The functional derivative $G(x,t;\Omega,\tau) = \delta X(x,t)/\delta\, \mathbf{do}_{\Omega,\tau}[\mathcal{F}]$, quantifying how the intervention affects location x at time t.
- **Causal flux** $\Phi_{\Omega\to\Gamma}$: The integrated influence from intervention region Ω to target region Γ: $\Phi_{\Omega\to\Gamma} = \int_\Gamma \int_\tau^\infty G(x,t;\Omega,\tau)\, dt\, dx$.
- **Causal cone**: The support of G: the set of (x,t) for which $G(x,t;\Omega,\tau) \neq 0$, delineating the spatio-temporal region of causal influence.