Equipping Graphical Models with Interventions and Interactions Simultaneously

James Enouen [1] **Angela Zhou** [1] **Yan Liu** [1]

Abstract

In this work, we reinvestigate the classical Markov equivalence classes (MECs) and interventional Markov equivalence classes ($\mathcal{I}$-MECs) with a new lens using higher-order feature interactions. We find that this perspective is particularly insightful for understanding statistical aspects (finite sample complexity) of recovering the true DAG, highlighting the shortcomings which must be faced in practical settings with finite sample availability. We propose this research direction can help close the gap between theoretical results on $\mathcal{I}$-MECs and practical approaches in Bayesian experiment design, serving as a possible theoretical support for results occurring in actual experiment data.

1. Introduction

The problem of Causal Discovery or Causal Structure Learning has had a long history of study (Pearl, 2009). Using directed graphical models and understanding their observational structure mark some of the most salient historical developments (Verma & Pearl, 1990). Adding interventional data has also proven key for understanding what can be recovered in the presence of experiment data (Eberhardt, 2007; Hauser & Bühlmann, 2012).

Across decades of research in causality, there have been vast improvements in the understanding of causal structure and significant advancements for structure learning algorithms. Nevertheless, sample complexity often remains a secondary consideration, being studied under various further assumptions (Kalisch & Bühlman, 2007; Gao et al., 2022). In this work, we show how incorporating higher-order interactions and unbounded interventions simultaneously paves a natural path for further studying these statistical considerations for causal discovery, especially in interventional settings.

*Equal contribution [1]University of Southern California, Los Angeles, California, USA. Correspondence to: James Enouen <enouen@usc.edu>.

Nonarchival Manuscript for the Workshop on Scaling Up Intervention Models at the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

2. Related Work

We review some of the main related works for structure learning *with interventions*. We see these works as often divided into two main directions, the theoretical and the practical, often done in parallel to one another.

Infinite Sample Identifiability The first is focusing on the characterization and identification of Markov equivalence classes (MECs) of graphical models under interventions (Hauser & Bühlmann, 2012; Yang et al., 2018). Much progress has been made on understanding the structure and other aspects like the worst case identifiability of structures (Eberhardt, 2007) and the combinatorial structure (Squires & Uhler, 2022) have also been explored.

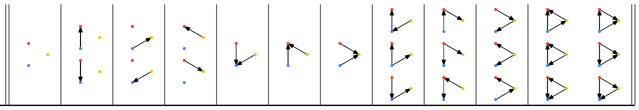
Finite Sample Bayesian Methods On the other side, there are many works doing Bayesian Optimal Experiment Design (Lindley, 1956) in the context of causal discovery, which focuses on achieving the best recovery of the graph under a Bayesian framework (Ghassami et al., 2018; Agrawal et al., 2019; Tigas et al., 2022). These mainly focus on improving the speed or efficiency of Bayes methods.

In addition, there is also (Acharya et al., 2018) looking at identification with few samples even under confounding and (Shiragur et al., 2024) reducing the number of CI tests needed. Another work inspired by similar considerations is (Kocaoglu, 2023), but it restricts itself to observational data and ultimately chooses a slightly different definition.

3. Problem Formulation

The classical case of $\mathcal{I}$-MECs (Hauser & Bühlmann, 2012) deals with choosing a set of interventions via $\mathcal{I} \subseteq \mathcal{P}([d])$ which can equally be written as a function $\mathcal{N} : \mathcal{P}([d]) \to \{0, \infty\}$. Whenever $I \in \mathcal{I}$ or equivalently $\mathcal{N}(I) = \infty$, we say that we have access to samples under intervention I. We use this notation to emphasize the fact that, implicitly, all of their theoretical results assume the population limit of infinite samples for each I s.t. $\mathcal{N}(I) = \infty$. In this work, we instead focus on the case of finite samples and accordingly use the function $\mathcal{N} : \mathcal{P}([d]) \to \mathbb{N}_0$ to track the number of samples n_I which have been collected under each intervention distribution, indexed by the $I \in \mathcal{P}([d])$.

For the simplicity of our theoretical results, we restrict to



$\mathcal{O}(t^2)$	$X_1 \perp X_2$	✓	✓	✓		✓						
	$X_1 \perp X_3$	✓		✓	✓			✓				
	$X_2 \perp X_3$	✓	✓		✓		✓					
$\mathcal{O}(t^3)$	$X_1 \perp X_2 \mid X_3$	✓	✓	✓					✓			
	$X_1 \perp X_3 \mid X_2$	✓		✓	✓						✓	
	$X_2 \perp X_3 \mid X_1$	✓	✓		✓					✓		
$\mathcal{O}(t^3)$	$X_1 \perp X_2, X_3$	✓		✓								
	$X_2 \perp X_1, X_3$	✓	✓									
	$X_3 \perp X_1, X_2$	✓			✓							

Table 1. A full table depicting the 9 Conditional Independence Tests (CITs) satisfied by each possible choice of the 11 Markov Equivalence Classes (MECs) for DAGs on three vertices. Each column represents a MEC and all DAGs within the class are depicted in the top row.

the regime where each variable's finite domain (t) is much larger than the total number of variables (d). We write this assumption as $d \ll t$. This enables us to refine the traditional notions of MEC and $\mathcal{I}$-MEC to versions which have a dependence on the finite number of available samples.

In particular, writing $n_I = \mathcal{N}(I)$, we may be concerned with whether or not there are enough samples to learn a k-dimensional function of the variables. In line with the curse of dimensionality, we know that the number of samples must be of at least size $\mathcal{O}(t^k)$. Accordingly, we will write that for each n_I, we calculate the 'order' of the number of samples we have collected with respect to our domain size t, and write k_I to be the largest integer such that $n_I \in \Omega(t^{k_I})$. Correspondingly, we write the function $\mathcal{K} : \mathcal{P}([d]) \to \mathbb{N}_0$.

4. Identifiability of $(\mathcal{I}, \mathcal{N})$-MECs

4.1. Lists of Independence Tests

We first recall that each DAG will lead to a list of conditional independences obeyed by and disobeyed by the induced distribution. Each MEC is then the set of all DAGs which obey the exact same conditional dependencies. In Table 1, we depict all 9 CITs for the 25 DAGs (11 MECs) on three variables. As d grows, it is difficult to exhaustively list all DAGs, all MECs, and all CITs as the number of DAGs, MECs, and CITs grow very rapidly in the number of variables. Taking d=4 leads to (543, 185, 55) and taking d=5 leads to (29281, 8782, 285), respectively, for instance.

Consequently, most existing algorithms instead directly choose tests which lead to the easiest recovery of graphical properties. However, this overlooks the sample size requirements inherited by their CIT choices.

4.2. Power of an Independence Test

In particular, if we would like to test for "$X_1 \perp X_2$", a fundamentally '2-dimensional' test requiring $\mathcal{O}(t^2)$ samples, while simultaneously wanting to test for "$X_1 \perp X_2 \mid X_3$", a fundamentally '3-dimensional' test requiring $\mathcal{O}(t^3)$ samples. Practically speaking, however, we may only have $\mathcal{O}(t^2)$ samples, meaning we can only perform the 2-dimensional test and our 3-dimensional test will have very low statistical power.

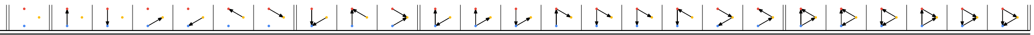
This 'dimensionality' interpretation is supported by our $d \ll t$ assumption and directly determines which CITs have statistical power. In particular, we will say that all k'-dimensional tests on the intervened distribution $p_I(x)$ where $k' \leq k_I := \mathcal{K}(I)$ are valid, and all k'-dimensional tests where $k' > k_I$ are underpowered and hence invalid.

In the observational case, this means that instead of the complete MECs, we would only have a coarser picture of the equivalence classes built by the subset of CITs which have been deemed valid. In Table 1, we can see how for $n \in \mathcal{O}(t^2)$, we would only have 8 equivalence classes, with the last 4 of 11 MECs being merged into a single class (containing the 15 DAGs in the top-right). These all disobey the three 2-dimensional independence tests: $X_1 \not\perp X_2$, $X_1 \not\perp X_3$, and $X_2 \not\perp X_3$.

4.3. Incorporating Interventional Data

Of course, we would also like to incorporate the data we have collected for each intervention $I \in \mathcal{P}([d])$. We will directly extend our previous arguments for CITs on the intervened distributions $p_I(x)$ for each $I \in \mathcal{P}([d])$. In Table 2, we depict all 9 CITs for the observational distribution and all 9 CITs for the $\{1\}$-interventional distribution. Because it is possible to identify beyond the MEC, we give each of the 25 DAGs their own column.

We use the potential outcomes or SWIG notation (Robins, 1986; Richardson & Robins, 2013) of $X_2(x_1)$ to help distinguish it from the observational case. Accordingly, one must take slight caution when interpreting x_1 or $X_2(x_1)$, because when we later take some CIT we must have a distribution over x_1, implying a choice of distribution over x_1. Although the actual choice of distribution may affect the statistical efficiency, we restrict our exploration to the static setting so it is sufficient for us to assume the uniform distribution.



$\emptyset \mapsto \mathcal{O}(t^2)$	$X_1 \perp\!\!\!\perp X_2$	✓	✓	✓	✓	✓			✓																	
	$X_1 \perp\!\!\!\perp X_3$	✓			✓	✓	✓	✓			✓															
	$X_2 \perp\!\!\!\perp X_3$	✓	✓	✓			✓	✓		✓																
$\emptyset \mapsto \mathcal{O}(t^3)$	$X_1 \perp\!\!\!\perp X_2 \mid X_3$	✓	✓	✓	✓	✓						✓	✓	✓												
	$X_1 \perp\!\!\!\perp X_3 \mid X_2$	✓			✓	✓	✓	✓										✓	✓	✓						
	$X_2 \perp\!\!\!\perp X_3 \mid X_1$	✓	✓	✓			✓	✓							✓	✓	✓									
$\emptyset \mapsto \mathcal{O}(t^3)$	$X_1 \perp\!\!\!\perp X_2, X_3$	✓			✓	✓																				
	$X_2 \perp\!\!\!\perp X_1, X_3$	✓	✓	✓																						
	$X_3 \perp\!\!\!\perp X_1, X_2$	✓					✓	✓																		
$\{1\} \mapsto \mathcal{O}(t^2)$	$x_1 \perp\!\!\!\perp X_2(x_1)$	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓				✓		✓	✓	✓	✓	✓			
	$x_1 \perp\!\!\!\perp X_3(x_1)$	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓		✓				✓	✓	✓	✓		✓		
	$X_2(x_1) \perp\!\!\!\perp X_3(x_1)$	✓	✓	✓			✓	✓		✓					✓		✓									
$\{1\} \mapsto \mathcal{O}(t^3)$	$x_1 \perp\!\!\!\perp X_2(x_1) \mid X_3(x_1)$	✓	✓	✓	✓	✓	✓			✓		✓	✓	✓			✓		✓	✓	✓	✓				
	$x_1 \perp\!\!\!\perp X_3(x_1) \mid X_2(x_1)$	✓	✓		✓	✓	✓	✓		✓		✓	✓		✓			✓	✓	✓	✓	✓				
	$X_2(x_1) \perp\!\!\!\perp X_3(x_1) \mid x_1$	✓	✓	✓			✓	✓		✓					✓	✓	✓									
$\{1\} \mapsto \mathcal{O}(t^3)$	$x_1 \perp\!\!\!\perp X_2(x_1), X_3(x_1)$	✓	✓		✓	✓	✓			✓		✓	✓						✓	✓	✓	✓				
	$X_2(x_1) \perp\!\!\!\perp x_1, X_3(x_1)$	✓	✓	✓			✓			✓							✓									
	$X_3(x_1) \perp\!\!\!\perp x_1, X_2(x_1)$	✓	✓				✓	✓		✓					✓											

Table 2. All 25 DAGs on three variables along with their complete set of conditional independence tests for both the observational setting ($I = \emptyset$) and the interventional setting while intervening on X_1 ($I = \{1\}$). Additionally organized by their MEC and the order of each conditional independence test.

Recall that we cannot identify the full MEC with only quadratic samples in the observational case. Surprisingly, we can in the interventional case by taking $n_\emptyset, n_{\{1\}}, n_{\{2\}}, n_{\{3\}}$ all in $\mathcal{O}(t^2)$ (moreover we can identify the DAG exactly). This demonstrates that the folk knowledge that interventions should be statistically more efficient for identifying DAG structure can be rigorously shown under the framework we introduce.

This additionally contrasts with the large body of work in interventional discovery which assumes the MEC as the starting point. Indeed, unless the cost of collecting interventional data is orders of magnitude more difficult than collecting observational data, it seems rather bizarre to assume starting knowledge of the MEC. This is perhaps antithetical to studying interventional settings in the first place and perhaps deserves greater scrutiny. In three variables, the quadratically identifiable MEC is still relatively informative; however, this only worsens as d grows.

5. Extension to Hypergraphs and Interactions

We now extend to higher-order interactions in the context of sample complexity concerns (Enouen & Sugiyama, 2024). Recent work has extended the classical MEC over DAGs to a hyper MEC (HMEC) over hyper DAGs (HDAGs) (Enouen et al., 2025). We reintroduce their main points herein.

Definition 5.1. *Conditional Multi-Independence Test.* We will write that three variables X_1, X_2, X_3 are multi-independent, denoted as $\perp\!\!\!\perp_3 (X_1, X_2, X_3)$ if we may write the distribution over the variables in the following form:

$$\log p(x_1, x_2, x_3) = f_{12}(x_1, x_2) + f_{13}(x_1, x_3) + f_{23}(x_2, x_3)$$

This extends on the typical notion of independence from an additive model perspective:

$$\log p(x_1, x_2) = f_1(x_1) + f_2(x_2)$$

Extensions to k-multi-dependence for $k > 3$ and for conditional multi-dependence are defined in the obvious way.

It is fairly straightforward to see that these tests are also identifiable directly from the observed distribution, and it can be further seen that these CMITs extend beyond the typical CITs which defined Markov equivalence. Accordingly, it is possible for them to define the hyper MEC via the equivalence over this larger set of CMITs.

5.1. Interactions using Hypergraphs

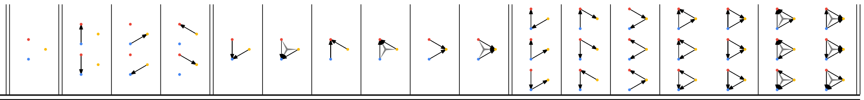
Recall that for a directed graph $\mathcal{G} = (V, E)$, we write $\text{Pa}_\mathcal{G}(j) = \{k \in [d] : (k, j) \in \mathcal{G}\}$ for the parents of a node. For a directed hypergraph $\mathcal{H} = (V, H)$ with $H \subseteq \{(S, j) : j \in V, S \subseteq (V - j)\}$, we write the hyperparents as $\text{HypPa}_\mathcal{H}(j) = \{S : (S, j) \in \mathcal{H}\}$. We may write the Markov and hyper-Markov properties as:

$$\log p(\boldsymbol{x}) = \sum_{i=1}^{d} \Big(\theta(x_i; \boldsymbol{x}_{\text{Pa}(i)}) - \mathcal{Z}(\boldsymbol{x}_{\text{Pa}(i)})\Big) \quad (1)$$

$$\log p(\boldsymbol{x}) = \sum_{i=1}^{d} \sum_{S \in \text{HypPa}(i)} \theta(x_i; \boldsymbol{x}_S) - \mathcal{Z}(\boldsymbol{x}_{\text{Pa}(i)}) \quad (2)$$

Theorem 5.2. *(Enouen et al. (2025)) The HMEC is the set of all HDAGs which have the same 'body' and same 'multi-colliders', extending the notion of skeleton and colliders in the DAG case.*

In the same way that a skeleton's edges describes a dependency which will never disappear under any conditioning set, a body's hyperedges describes a multidependency which will never disappear under any conditioning set. In the same way that a collider describes two parents whose dependency is induced by their joint child, a multi-collider describes three or more parents whose dependency is induced by their joint child.



$\mathcal{O}(t^2)$	$X_1 \perp\!\!\!\perp X_2$	✓	✓	✓		✓	✓									
	$X_1 \perp\!\!\!\perp X_3$	✓		✓	✓					✓	✓					
	$X_2 \perp\!\!\!\perp X_3$	✓	✓		✓			✓	✓							
$\mathcal{O}(t^3)$	$X_1 \perp\!\!\!\perp X_2 \mid X_3$	✓	✓	✓								✓				
	$X_1 \perp\!\!\!\perp X_3 \mid X_2$	✓		✓	✓									✓		
	$X_2 \perp\!\!\!\perp X_3 \mid X_1$	✓	✓		✓								✓			
$\mathcal{O}(t^3)$	$X_1 \perp\!\!\!\perp X_2 X_3$	✓		✓												
	$X_2 \perp\!\!\!\perp X_1 X_3$	✓	✓													
	$X_3 \perp\!\!\!\perp X_1 X_2$	✓			✓											
$\mathcal{O}(t^3)$	$\perp\!\!\!\perp_3 (X_1, X_2, X_3)$	✓	✓	✓	✓	✓		✓		✓		✓	✓	✓	✓	

Table 3. All 34 HDAGs depicted in each of the 15 HMECs. The 3D multidependence test (CMIT) which distinguishes the higher-order structure is also included in addition to the 9 CITs.

5.2. Interactions and Interventions with $(\mathcal{I}, \mathcal{N})$-HMECs

Based on the discussion in Section 4, it should be straightforward to first see how we can extend their notion of HMEC to the coarser version which is identifiable in finite samples. For example, any $\perp\!\!\!\perp_3$-CMIT testing requires a minimum of three dimensions and thus requires at least $\mathcal{O}(t^3)$ samples.

We depict in Table 3 the extension of Table 1 to the case of HDAGs and HMECs. Once again, we see a similar partitioning where the full observational structure is not identified in only quadratic samples. Although we only depict the small case of d=3, it is already enough to show the differences without infinite samples. Moreover, the impact of our finite sample focus only has a greater and greater impact as we increase to larger d. Once again, the interventional data can be incorporated and proves critical for identifying structure.

6. Relevance to Causal Inference

Another key area of impact, especially since the hypergraph extension blends graphical structure with semiparametric regression model specification, is causal estimation. Interventions can improve causal estimation in complex settings, even without interventions on the final treatment of interest.

6.1. Statistical Efficiency

The area of causal estimation often assumes unconfoundedness and uses backdoor adjustments (e.g. via AIPW). Although this overlooks graphical complexity, it enables the literature to dive deeper into subtleties on estimation and efficiency. The usage of sieve methods (Newey, 1997; Chen, 2007) allows for a gradually expanding structure based on n, mirroring the growth in complexity in $(\mathcal{I}, \mathcal{N})$-MECs. Moreover, some work highlights the additive interaction model for favorable finite sample complexity scaling for real-world settings (Andrews & Whang, 1990).

Moreover, in the case of double machine learning (DML), having robust convergence rates under partial misspecification is a key topic of interest (Chernozhukov et al., 2018). It is imagined that these concerns can be revisited in a hypergraphical model and moreover that the hypergraphical model can automatically specify necessary interactions as discussed in the next section.

6.2. Valid Adjustments and Bad Controls

Even for simple backdoor identification, graphical structure informs estimation implications (via bias or precision), (Cinelli et al., 2024), and gives tools for finding valid adjustment sets (Textor et al., 2017).

Beyond valid adjustment sets, recent work further studies efficiency thereof (Rotnitzky & Smucler, 2020). Similar extensions to the selection of valid adjustment *variables* would be the selection of valid adjustment *interactions*, closely related to regression model selections like additive interaction models. With partial knowledge on the *HDAG*, if analogous adjustment tools could be developed, it too can be leveraged directly for the causal inference practitioner.

7. Conclusion

The biggest takeaway we hope can be that the MEC is not a powerful enough description to describe the statistical nuance in recovering the causal graphical model. Moreover, the language of interventions as well as interactions provides an extremely powerful first look at the statistical shortcomings the classical MEC and $\mathcal{I}$-MEC approaches. It is hoped that for future works on experimental design, these statistical aspects of recovery can be more directly incorporated to use reasoning under uncertainty while also shedding the dependency on the MEC starting point. Further developments and refinements of the statistical theory presented herein would likely continue to lead to downstream benefits.

References

Acharya, J., Bhattacharyya, A., Daskalakis, C., and Kandasamy, S. Learning and testing causal models with interventions. *Advances in Neural Information Processing Systems*, 31, 2018.

Agrawal, R., Squires, C., Yang, K., Shanmugam, K., and Uhler, C. Abcd-strategy: Budgeted experimental design for targeted causal structure discovery. In Chaudhuri, K. and Sugiyama, M. (eds.), *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 3400–3409. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/agrawal19b.html.

Andrews, D. W. K. and Whang, Y.-J. Additive interactive regression models: Circumvention of the curse of dimensionality. *Econometric Theory*, 6(4):466–479, 1990. ISSN 02664666, 14694360. URL http://www.jstor.org/stable/3532092.

Chen, X. Chapter 76 large sample sieve estimation of semi-nonparametric models. volume 6 of *Handbook of Econometrics*, pp. 5549–5632. Elsevier, 2007. doi: https://doi.org/10.1016/S1573-4412(07)06076-X. URL https://www.sciencedirect.com/science/article/pii/S157344120706076X.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018. ISSN 1368-4221. doi: 10.1111/ectj.12097. URL https://doi.org/10.1111/ectj.12097.

Cinelli, C., Forney, A., and Pearl, J. A crash course in good and bad controls. *Sociological Methods & Research*, 53(3):1071–1104, 2024. doi: 10.1177/00491241221099552. URL https://doi.org/10.1177/00491241221099552.

Eberhardt, F. *Causation and Intervention*. PhD thesis, Carnegie Mellon University, 2007.

Enouen, J. and Sugiyama, M. A complete decomposition of kl error using refined information and mode interaction selection, 2024. URL https://arxiv.org/abs/2410.11964.

Enouen, J., Zheng, Y., Ng, I., Liu, Y., and Zhang, K. Higher-order causal structure learning with additive models. *Under Review*, 2025.

Gao, M., Tai, W. M., and Aragam, B. Optimal estimation of gaussian dag models. In *International Conference on Artificial Intelligence and Statistics*, pp. 8738–8757. PMLR, 2022.

Ghassami, A., Salehkaleybar, S., Kiyavash, N., and Bareinboim, E. Budgeted experiment design for causal structure learning. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1724–1733. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/ghassami18a.html.

Hauser, A. and Bühlmann, P. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13(79):2409–2464, 2012. URL http://jmlr.org/papers/v13/hauser12a.html.

Kalisch, M. and Bühlman, P. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, 8(3), 2007.

Kocaoglu, M. Characterization and learning of causal graphs with small conditioning sets. *Advances in Neural Information Processing Systems*, 36:74140–74179, 2023.

Lindley, D. On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27:986–1005, 1956. URL https://api.semanticscholar.org/CorpusID:123582195.

Newey, W. K. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168, 1997. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(97)00011-0. URL https://www.sciencedirect.com/science/article/pii/S0304407697000110.

Pearl, J. *Causality*. Cambridge University Press, 2 edition, 2009.

Richardson, T. S. and Robins, J. M. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013, 2013.

Robins, J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512, 1986.

Rotnitzky, A. and Smucler, E. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *Journal of Machine Learning Research*, 21(188):1–86, 2020. URL http://jmlr.org/papers/v21/19-1026.html.

Shiragur, K., Zhang, J., and Uhler, C. Causal discovery with fewer conditional independence tests. *arXiv preprint arXiv:2406.01823*, 2024.

Squires, C. and Uhler, C. Causal structure learning: A combinatorial perspective. *Foundations of Computational Mathematics*, 23(5):1781–1815, August 2022. ISSN 1615-3383. doi: 10.1007/s10208-022-09581-9. URL http://dx.doi.org/10.1007/s10208-022-09581-9.

Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., and Ellison, G. T. Robust causal inference using directed acyclic graphs: the r package 'dagitty'. *International Journal of Epidemiology*, 45(6):1887–1894, 01 2017. doi: 10.1093/ije/dyw341. URL https://doi.org/10.1093/ije/dyw341.

Tigas, P., Annadani, Y., Jesson, A., Schölkopf, B., Gal, Y., and Bauer, S. Interventions, where and how? experimental design for causal models at scale. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=ST5ZUlz_3w.

Verma, T. and Pearl, J. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '90, pp. 255–270, USA, 1990. Elsevier Science Inc. ISBN 0444892648.

Yang, K., Katcoff, A., and Uhler, C. Characterizing and learning equivalence classes of causal DAGs under interventions. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5541–5550. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/yang18a.html.

A. Additional Figures

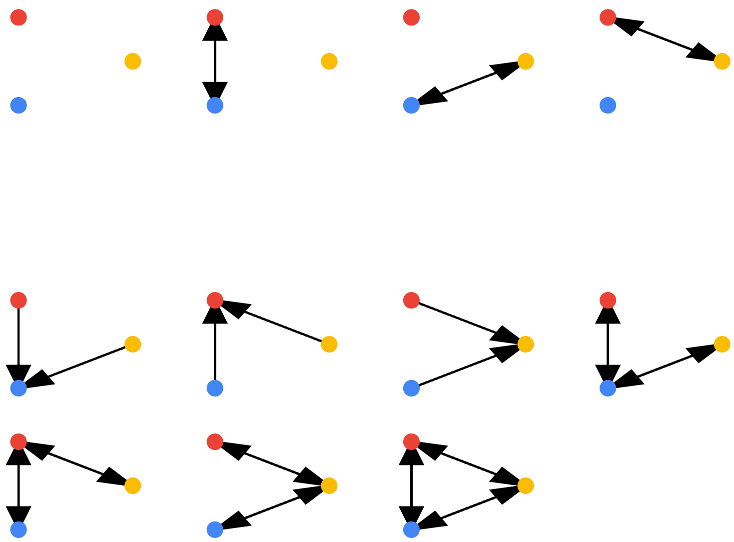


Figure 1. All 3D MECs

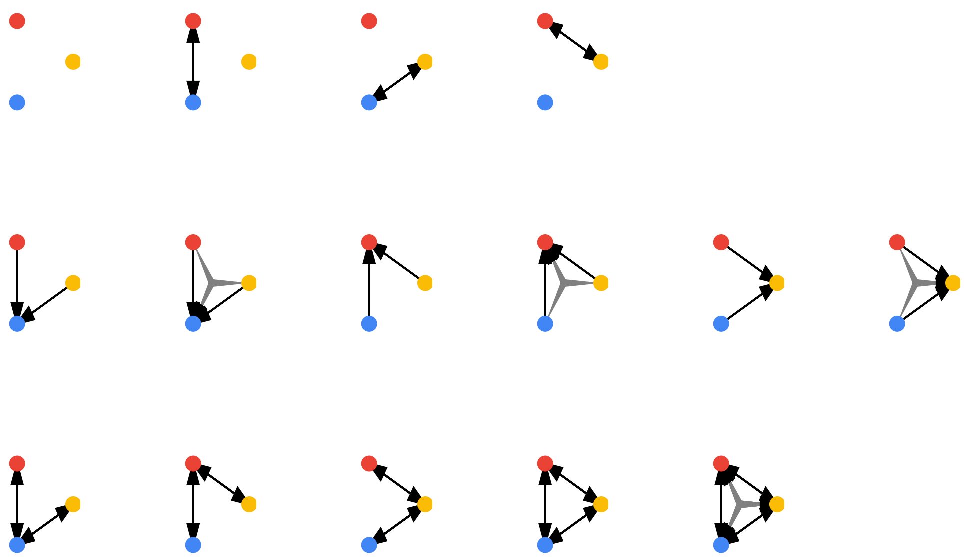


Figure 2. All 3D HMECs

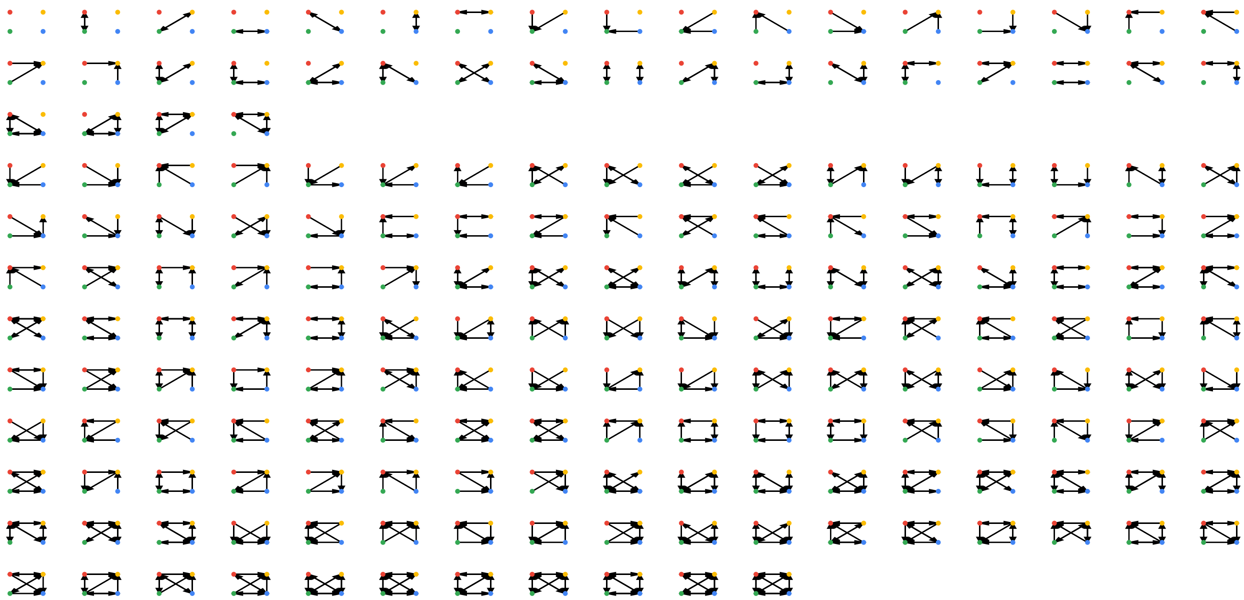


Figure 3. All 4D MECs

Figure 4. All 4D HMECs