Diverse Construction of Sidon Sets with Generative Flow Networks

Massimiliano Esposito [*1] **Dmitry Zubarev** [*2] **Besart Shyti** [3] **Alberto Alfarano** [1]

Abstract

Sidon sets — subsets of $[n]$ with all pairwise sums distinct — admit many structurally different maximum-size solutions, making the family $\mathcal{M}_n$ of maximum Sidon sets in $[n]$ a clean benchmark for diverse combinatorial construction under a sharp global constraint. We cast maximum-Sidon-set construction as a sequential GFlowNet with ordered trajectories and exact feasibility masking, and study three ingredients: a number-theoretic transformer encoding modular, Fourier, and sum-conflict structure; an *adherence loss* over a partial catalogue of maximum sets at the target; and *behavioural cloning* transferring from sizes $n < n^\star$ to an unseen $n^\star$. At $n \in \{30, 50\}$ the number-theoretic backbone consistently dominates MLP and vanilla-transformer baselines, isolating inductive bias as the dominant source of gain. At $n = 80$ scratch training collapses; BC-initialized fine-tuning with target-size adherence supervision reaches total catalogue coverage $\approx$ 0.82–0.83, with a held-out discovery rate of 0.12–0.21. A $30 \to \cdots \to 80$ curriculum reaches 0.77 without using the $n = 80$ catalogue. We therefore present the method as a controlled preliminary step toward scalable AI-assisted mathematical construction, rather than as a demonstration of autonomous discovery at non-enumerable scales.

1. Introduction

Recent AI systems have shown that learning-based search can discover strong discrete constructions. AlphaTensor, for example, used reinforcement learning to discover provably correct matrix multiplication algorithms, while FunSearch found improved constructions for the cap set problem through program search with large language models (**??**).

However, many problems in combinatorics are not naturally judged only by one best object. Instead, they involve a family of feasible constructions, often with many structurally different high-quality solutions. Sidon sets are a clean example of this regime. In this paper, we study Sidon subsets of $[n] = \{1, \ldots, n\}$, namely subsets whose pairwise sums are all distinct. More broadly, Sidon sets are classical objects in additive combinatorics, with a rich literature on their structure and extremal size (**??**).

This viewpoint makes Sidon-set construction a natural test bed for generative combinatorial optimization. GFlowNets are designed to construct compositional objects through stochastic trajectories so that terminal objects are sampled approximately in proportion to reward, which makes them appealing when one wants diverse high-reward candidates rather than only a single optimum (**??**). Prior work has also improved GFlowNet training through trajectory-level objectives and shown that GFlowNets can amortize search on combinatorial optimization tasks while preserving diversity across solutions (**??**).

In this work, we ask whether a GFlowNet can learn to construct Sidon sets. We formulate the problem as a sequential decision process in which states are partial Sidon sets and actions either add a feasible new element or terminate the construction. Our formulation uses ordered trajectories to remove permutation symmetries, exact validity masking to enforce the Sidon constraint during generation, and a simple backward process induced by the construction order.

Our main empirical finding is that this domain is difficult for GFlowNets trained from scratch. At $n = 80$, a standalone GFlowNet fails to adequately cover the family of maximum Sidon sets under our evaluation protocol. In contrast, behavioural cloning pretraining followed by bridged GFlowNet fine-tuning with target-size adherence supervision achieves total catalogue coverage of approximately 0.82–0.83 of $\mathcal{M}_{80}$, improving both solution quality and diversity. However, held-out discovery remains substantially lower (0.12–0.21 in the strongest BC-initialized runs), indicating that the headline coverage reflects both recovery of supervised catalogue elements and partial generalization beyond them. These results suggest that imitation can provide a strong inductive bias for GFlowNet training in sparse

[*]Equal contribution [1]AxiomMath [2]IBM [3]Independent Researcher. Correspondence to: Massimiliano Esposito <max@axiommath.ai>, Dmitry Zubarev <dmitry.zubarev@ibm.com>, Alberto Alfarano <alberto@axiommath.ai>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

combinatorial domains, and that Sidon sets offer a useful benchmark for studying hybrid generative-search methods.

Scope. The present paper is a controlled catalogue-based study rather than a demonstration of fully autonomous mathematical discovery. All quantitative experiments are in regimes where exact reference catalogues are available, and the strongest $n = 80$ result uses a substantial target-size adherence pool. We use these enumerable settings because they allow exact measurement of total coverage, held-out discovery, and overfitting to the adherence pool; extending the method to genuinely non-enumerable regimes remains future work.

2. Background and problem setting

Sidon sets are classical objects in additive combinatorics. Historically, they arose from a question of Sidon on trigonometric series, and the area has since developed connections to coding theory and communication/synchronization problems (**??**).

For $n \in \mathbb{N}$, let $[n] := \{1, \dots, n\}$. A set $S \subseteq [n]$ is a *Sidon set* if

$$a + b = c + d \quad \Longrightarrow \quad \{a, b\} = \{c, d\}$$

for all $a, b, c, d \in S$. Equivalently, all sums $a + b$ with $a, b \in S$ and $a \leq b$ are distinct. For example, $\{1, 4, 9\} \subseteq [9]$ is a Sidon set, whereas $\{1, 2, 3\}$ is not because $1+3 = 2+2$. A larger example in $[9]$ is $\{1, 2, 4, 8\}$, which is also Sidon. Since every subset of a Sidon set is again Sidon, the space of valid constructions is downward closed and typically contains many solutions of different sizes.

We write

$$M(n) := \max\{|S| : S \subseteq [n],\ S \text{ is a Sidon set}\}$$

for the maximum cardinality of a Sidon subset of $[n]$, and

$$\mathcal{M}_n := \{S \subseteq [n] : S \text{ is Sidon and } |S| = M(n)\}$$

for the family of *maximum Sidon sets*. Classical results imply that $M(n) = (1 + o(1))\sqrt{n}$ (**?**), while an explicit upper bound is

$$M(n) < n^{1/2} + n^{1/4} + \tfrac{1}{2}$$

(**?**). In this work, the goal is not only to find one large Sidon set, but to discover as many distinct elements of $\mathcal{M}_n$ as possible.

Evaluation metrics. Let $p_\theta(S)$ denote the terminal probability assigned by the trained GFlowNet to a terminal Sidon set S. Given a sampling budget B, let $\mathcal{D}_\theta^{(B)}$ be the set of distinct terminal Sidon sets obtained from B model samples. We report the following main metrics.

Coverage fraction.

$$\mathrm{CovFrac}_B(\theta) = \frac{|\mathcal{D}_\theta^{(B)} \cap \mathcal{M}_n|}{|\mathcal{M}_n|}.$$

Normalized entropy on maximum solutions. Let

$$Z_\theta^{\max} := \sum_{S \in \mathcal{M}_n} p_\theta(S), \qquad \tilde{p}_\theta(S) := \frac{p_\theta(S)}{Z_\theta^{\max}} \quad \text{for } S \in \mathcal{M}_n.$$

We then define

$$H_{\max}(\theta) = -\frac{1}{\log |\mathcal{M}_n|} \sum_{S \in \mathcal{M}_n} \tilde{p}_\theta(S) \log \tilde{p}_\theta(S).$$

This metric lies in $[0, 1]$ and equals 1 when the model distributes its mass uniformly over maximum Sidon sets.

Reference families and held-out evaluation. The purpose of this split is to separate maximum solutions used to provide supervision from maximum solutions used only for evaluation. When the full family $\mathcal{M}_n$ is exactly known, this distinction can be made directly inside $\mathcal{M}_n$. In larger regimes, where exhaustive enumeration of $\mathcal{M}_n$ may be infeasible, the same idea can still be applied to a nonempty reference family $R_n \subseteq \mathcal{M}_n$ obtained from partial enumeration, cached search, or previously discovered maximum solutions. Families $\mathcal{M}_{n'}$ for smaller values $n' < n$ can also serve as source families for behavioural cloning or curriculum learning, providing an inductive bias before evaluation at the larger target n.

Known-solution split. For values of n where $\mathcal{M}_n$ is known exactly, we partition it as

$$\mathcal{M}_n = A_n \sqcup H_n,$$

where A_n is an adherence set used during training and H_n is a held-out set used only for evaluation. This lets us distinguish memorization of known maximum solutions from generalization to unseen ones. More generally, when only a partial reference family $R_n \subseteq \mathcal{M}_n$ is available, we use the analogous split

$$R_n = A_n \sqcup H_n,$$

and interpret the held-out metrics as generalization within the available reference catalogue rather than coverage of the entire unknown family $\mathcal{M}_n$. Consequently, in the target-size adherence experiments at $n = 80$, the adherence pool should be viewed as substantial catalogue supervision rather than evidence of catalogue-free discovery.

3. GFlowNet formulation

We cast Sidon-set construction as a sequential decision process over partial valid sets. A state is an ordered partial Sidon set

$$s = (x_1, \dots, x_k), \qquad 1 \le x_1 < \cdots < x_k \le n,$$

such that the underlying set $S(s) := \{x_1, \dots, x_k\}$ is Sidon. The initial state is the empty state $s_0 = \emptyset$.

For a state $s = (x_1, \dots, x_k)$, define the set of already realized pairwise sums

$$\Sigma(s) := \{x_i + x_j : 1 \le i \le j \le k\}.$$

A forward action is either STOP or the insertion of a new element $x \in [n]$. Under our ordered construction rule, an insertion is admissible only if $x > x_k$ (or $s = \emptyset$) and adding x preserves the Sidon property. Equivalently, x is feasible from s if

$$\big(\{x + x_i : 1 \le i \le k\} \cup \{2x\}\big) \cap \Sigma(s) = \emptyset.$$

We denote by $\mathcal{A}(s)$ the set of feasible forward actions together with STOP.

A complete trajectory therefore has the form

$$\emptyset \to (x_1) \to (x_1, x_2) \to \cdots \to (x_1, \dots, x_m) \to \text{STOP},$$

and terminates at the Sidon set $S = \{x_1, \dots, x_m\}$.

Ordered trajectories. The ordering constraint removes permutation symmetry: each terminal Sidon set

$$S = \{x_1 < \cdots < x_m\}$$

corresponds to a unique construction trajectory. This avoids representing the same object through $m!$ equivalent insertion orders and simplifies both training and evaluation.

Exact feasibility masking. At every state, we apply an exact mask over the forward action space and assign zero probability to invalid insertions. Concretely, logits corresponding to actions outside $\mathcal{A}(s)$ are set to $-\infty$ before normalization. As a result, every sampled prefix remains a valid partial Sidon set throughout generation. This is particularly important in the sparse regime considered here, where unmasked sampling would waste many rollouts on invalid actions.

Backward process. Because trajectories are ordered, every non-empty state has a unique predecessor:

$$(x_1, \dots, x_k) \mapsto (x_1, \dots, x_{k-1}).$$

We therefore use the deterministic backward process that removes the most recently inserted element. In particular, if $s' = (x_1, \dots, x_k)$ and $s = (x_1, \dots, x_{k-1})$, then

$$P_B(s \mid s') = 1.$$

This eliminates the need to learn a separate backward policy.

Trajectory-balance objective. Let $\tau = (s_0, a_0, \dots, s_T)$ be a complete trajectory ending at terminal Sidon set S, and let $R(S) > 0$ be the terminal reward. Since the backward process is deterministic, the trajectory-balance objective reduces to

$$\mathcal{L}_{\text{TB}}(\tau) = \left(\log Z_\theta + \sum_{t=0}^{T-1} \log P_{F,\theta}(a_t \mid s_t) - \log R(S)\right)^2,$$

where the sum runs over the forward actions in the trajectory, including the terminal STOP action. In the next section, we combine this objective with supervised components used for pretraining and, where applicable, adherence to known maximum Sidon sets (**???**).

4. Model and training

4.1. Practical state representation and policy architecture

We represent a state using three groups of features.

First, we encode membership of the current Sidon set with a binary indicator vector

$$b^{\text{set}}(s) \in \{0,1\}^n,$$

whose i-th entry indicates whether $i \in S(s)$.

Second, we encode the realized pairwise sums with a binary vector

$$b^{\text{sum}}(s) \in \{0,1\}^{2n-1},$$

indexed by the possible sums $2, \dots, 2n$, where the entry corresponding to u is 1 iff $u \in \Sigma(s)$. For example, if $S(s) = \{1, 4, 9\}$, then the realized sums are

$$\{2, 5, 8, 10, 13, 18\}.$$

Third, we add scalar and sequence-level features describing construction progress: the normalized ambient size, the current set size, the normalized last-inserted element (or 0 for the empty state), and the ordered list of currently selected elements padded to a fixed length. For per-element architectures we additionally compute, for every candidate $x \in \{1, \dots, n\}$, a vector of number-theoretic features — scale features (x/n, x^2/n^2, $\sqrt{x}/\sqrt{n}$, $\log x / \log n$), modular residues $x \bmod p$ and quadratic-residue indicators $\mathbf{1}[x \text{ is a QR mod } p]$ for an adaptively

chosen set of primes $p \leq n$, and Fourier features $\{\sin(2\pi kx/n), \cos(2\pi kx/n)\}_{k=1}^{K}$ that together span all periodic functions on $\{1, \ldots, n\}$ — together with two per-candidate dynamic features: a legality indicator and a sum-conflict count $c(x) = |\{y \in S(s) : x + y \in \Sigma(s)\}|$ encoding the Sidon constraint.

All architectures share this core state representation. We evaluate three encoders:

1) MLP: two hidden layers of size 512. The model consumes the concatenated bitstring and scalar features directly and outputs $n + 1$ logits.

2) Legacy Transformer encoder: a 2-layer transformer over a sequence of length 2, consisting of a state token and a learned CLS token, followed by an $n + 1$-dimensional output head.

3) Number-theoretic Transformer: a 4-layer pre-norm transformer of hidden dimension $d_{\text{model}} \in \{128, 256\}$ (depending on the experiment) with 8 attention heads. The model consumes the same bitstring and scalar features tokenized as one token per candidate $x \in \{1, \ldots, n\}$ plus a learned CLS token and outputs $n + 1$ logits.

Each model outputs one logit for every candidate insertion $x \in [n]$ together with one STOP logit. We then apply the exact feasibility mask from Section 3 before normalizing to obtain the forward policy $P_{F,\theta}(\cdot \mid s)$. When the reward is approximately constant on maximum solutions and comparatively small on submaximal ones, the corresponding normalizing constant is of the same order as the number of maximum sets. This motivates initializing $\log Z_\theta$ from $\log |\mathcal{M}_n|$ when a reference family is available, or from a nearby extrapolated value when only an approximate neighborhood is known. This is exactly our case since we can conjecture the cardinality of $\mathcal{M}_n$ after the first tabulations summarized in **??** (**??**). For some experiments we then continue optimizing $\log Z_\theta$ jointly with the policy during training.

4.2. GFlowNet objective

Reference family and split. For every target n we work with a reference family $\mathcal{R}_n$ of known maximum Sidon sets, equal to $\mathcal{M}_n$ when exact enumeration is tractable and to a partial precomputed cache of verified maximum sets otherwise. The adherence machinery introduced below depends only on having a non-empty $\mathcal{R}_n$, so it remains usable in larger-n regimes in which exhaustive enumeration is infeasible but a partial catalogue of known maximum Sidon sets is available — an arrangement common to many combinatorial discovery tasks. In such regimes $\mathcal{R}_n$ should be understood as a memory of verified constructions obtained from partial enumeration, heuristic search, previous model rollouts, or external mathematical knowledge, not as oracle access to the full target family.

We split $\mathcal{R}_n$ into an $80/20$ partition

$$\mathcal{R}_n \;=\; \mathcal{A}_n \sqcup \mathcal{H}_n,$$

where $\mathcal{A}_n$ feeds the adherence loss defined below and $\mathcal{H}_n$ is reserved strictly for held-out evaluation. This fixed split should be interpreted as a pragmatic preliminary protocol rather than an optimized choice. It can be stringent, especially when $|\mathcal{R}_n|$ is small or only partially enumerated, because the held-out subset may contain relatively few maximum solutions and the outcome can depend on the split seed. We nevertheless use it here because it preserves a large adherence pool for stable training while keeping a non-negligible subset of known maximum solutions completely absent from all training-time losses. Thus the held-out metrics provide a simple first test of whether the policy assigns probability beyond the specific solutions used by the adherence objective. In future work, this protocol can be strengthened by averaging over several split seeds or by varying the held-out fraction.

Trajectory-balance residual. For a complete trajectory $\tau = (s_0 \to \cdots \to s_K \to \text{STOP})$ with terminal Sidon set S, the trajectory-balance residual is

$$r(\tau) \;=\; \log Z_\theta + \log P_{F,\theta}(\tau) - \log R(S),$$

where $\log Z_\theta$ is a learnable scalar parameter and $\log P_{F,\theta}(\tau)$ is the sum of forward log-probabilities of the chosen actions, including the terminal STOP. We use either a clipped quadratic $\mathcal{L}_{\text{TB}}^{\text{l}^2}(r) = \text{clip}(r, \pm\delta)^2$ or a pseudo-Huber form $\mathcal{L}_{\text{TB}}^{\text{ph}}(r) = \delta^2\big(\sqrt{1 + (r/\delta)^2} - 1\big)$ with $\delta = 20$; the pseudo-Huber form avoids the gradient saturation of the clipped form on outlier residuals.

Terminal reward. In the *diversity regime* the goal is to sample $\mathcal{M}_n$ uniformly, so we use an indicator reward

$$R(S) = \begin{cases} 1 & |S| = M(n), \\ \varepsilon = 10^{-8} & |S| < M(n), \end{cases} \tag{1}$$

where $M(n)$ is the maximum Sidon-set cardinality on $[n]$ (**??**) and the floor ε only keeps $\log R$ finite for the dead-end penalty $\mathcal{L}_{\text{dead}}$ in (**??**). Because R is constant on $\mathcal{M}_n$, trajectory balance forces $P_{F,\theta}(\tau(S)) \propto 1/Z_\theta$ for every $S \in \mathcal{M}_n$, so $Z_\theta \to |\mathcal{M}_n|$ and the terminal distribution converges to the uniform on $\mathcal{M}_n$. The TB loss is computed only on successful trajectories ($|S| = M(n)$), so ε enters the objective only through $\mathcal{L}_{\text{dead}}$.

On-policy and adherence terms. The on-policy objective averages the residual loss over a minibatch $\mathcal{B}$ of policy rollouts:

$$\mathcal{L}_{\text{TB}}^{\text{sample}} = \frac{1}{|\mathcal{B}|} \sum_{\tau \in \mathcal{B}} \mathcal{L}_{\text{TB}}\big(r(\tau)\big).$$

To inject the supervisory signal contained in $\mathcal{A}_n$, observe that each $S = \{x_1 < \cdots < x_m\} \in \mathcal{A}_n$ determines the unique expert trajectory $\tau(S)\colon \emptyset \to (x_1) \to \cdots \to (x_1, \ldots, x_m) \to \textsc{Stop}$. The adherence loss is the mean residual loss along these expert trajectories,

$$\mathcal{L}_{\text{adh}} = \frac{1}{|\mathcal{A}_n|} \sum_{S \in \mathcal{A}_n} \mathcal{L}_{\text{TB}}\big(r(\tau(S))\big).$$

This term is our mechanism for exploiting a partial library of known maximum solutions: it remains well-defined whenever $\mathcal{R}_n$ is non-empty. It should not be interpreted as oracle access to the full target family in non-enumerable regimes. In those regimes the pool would be sparse, and the on-policy GFlowNet objective must explore beyond it. Our experiments test this mechanism in the controlled setting where $\mathcal{R}_n$ is known exactly or cached, so robustness under very sparse reference pools remains open.

Auxiliary losses. Two stabilizers improve the on-policy term in practice. A dead-end penalty

$$\mathcal{L}_{\text{dead}} = \frac{1}{|\mathcal{B}_{\text{dead}}|} \sum_{\tau \in \mathcal{B}_{\text{dead}}} \max\big(\log P_{F,\theta}(\tau),\, -50\big)$$

collects the rollouts in $\mathcal{B}_{\text{dead}} \subseteq \mathcal{B}$ that fail to reach a complete maximum Sidon set, with a floor at -50 that prevents runaway logits once the dead-end probability is already negligible. An entropy bonus

$$\bar{H} = \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\big[H\big(P_{F,\theta}(\cdot \mid s_t)\big)\big],$$

averaged over forward steps along the rollout, encourages stochastic action selection and discourages mode collapse.

Final GFlowNet objective. Combining all four ingredients, the GFlowNet objective at target n is

$$\mathcal{L}_n = \mathcal{L}_{\text{TB}}^{\text{sample}} + \beta_{\text{dead}}\, \mathcal{L}_{\text{dead}} - \beta_{\text{ent}}\, \bar{H} + \omega(t)\, \mathcal{L}_{\text{adh}}, \tag{2}$$

with scalar weights $\beta_{\text{dead}}, \beta_{\text{ent}} > 0$ and an annealing schedule $\omega(t) = \max\big(0,\, 1 - t/T_{\text{anneal}}\big)$ that linearly decays the adherence weight to zero over training. This same objective is used in both the scratch and BC-initialized settings; the only difference between the two is the initialization of the policy parameters, which we describe next.

4.3. Behavioural cloning pretraining and bridge

We consider behavioural cloning (BC) as an alternative training signal, either standalone or as pretraining for the GFlowNet objective. BC is implemented as next-token prediction on ordered expert trajectories drawn from a *multi-n pool* $\bigcup_k \mathcal{A}_{n_k}$, where the BC pretraining sizes $\{n_1, \ldots, n_K\}$ are typically smaller than the GFlowNet target $n^\star$ and are specified per experiment in **??**.

For $S = \{x_1 < \cdots < x_m\} \in \mathcal{A}_{n_k}$ inducing the expert trajectory $\tau(S)\colon \emptyset \to (x_1) \to \cdots \to (x_1, \ldots, x_m) \to \textsc{Stop}$, let $s_t = (x_1, \ldots, x_t)$ and

$$a_t^\star = \begin{cases} x_{t+1}, & t < m, \\ \textsc{Stop}, & t = m. \end{cases}$$

We minimize the masked next-token cross-entropy

$$\begin{aligned} \mathcal{L}_{\text{BC}} = &-\frac{1}{\sum_k |\mathcal{A}_{n_k}|\,(m_k + 1)} \\ &\times \sum_k \sum_{S \in \mathcal{A}_{n_k}} \sum_{t=0}^{m_k} \log P_{F,\theta}(a_t^\star \mid s_t), \end{aligned} \tag{3}$$

where the forward policy is normalized only over the feasible (masked) actions at each state and m_k is the size of the maximum Sidon sets at n_k. Writing the loss as a per-token average makes the BC signal comparable across pools of different sizes.

Bridge to GFlowNet fine-tuning. When BC is used as a prior for GFlowNet training at a target $n^\star$ (typically $n^\star > \max_k n_k$), we transiently retain the BC term over the first T_{bridge} steps of fine-tuning, so that the rollout-driven exploration of the GFlowNet phase does not abruptly overwrite the BC prior. The fine-tuning loss is

$$\begin{aligned} \mathcal{L}_t^{\text{bridge}} &= \mathcal{L}_{n^\star} + \lambda(t)\, \mathcal{L}_{\text{BC}}^{(n^\star)}, \\ \lambda(t) &= \lambda_0 \max\big(0,\, 1 - t/T_{\text{bridge}}\big). \end{aligned} \tag{4}$$

where $\mathcal{L}_{n^\star}$ is the GFlowNet objective (**??**) at the target and $\mathcal{L}_{\text{BC}}^{(n^\star)}$ is the BC loss (**??**) restricted to the target-$n^\star$ adherence pool $\mathcal{A}_{n^\star}$. For $t \geq T_{\text{bridge}}$ the BC term is fully removed and only (**??**) remains. This bridge variant lets the GFlowNet phase start from a strong supervised prior while still converging to the trajectory-balance solution.

5. Experimental setup

Our experiments are organized into two complementary scopes. The *GFlowNet scope* starts at the small problem sizes $n \in \{30, 50\}$, where we compare architectural inductive biases under matched training conditions within a 10^4-update budget, and then scales up to $n = 80$, where we additionally evaluate supplementary inductive

bias — a ground-truth log-partition initialization, an imitation prior, or a curriculum warm-up. The *behavioural-cloning scope* is multi-n by construction. We pretrain BC supervisedly on the ascending expert trajectories of $\mathcal{M}_n$ for $n \in \{39, 40, 44, 49, 50, 55, 64, 65\}$ and evaluate exclusively at the unseen target $n = 80$; the held-out test is therefore a strict generalization *across* problem sizes, not a within-problem split.

The two scopes are realised by five experimental families described in (i)–(v) below. All families share the optimization budget, the rollout batch size, the on-policy exploration schedule, and the $80/20$ adherence/held-out split (**??**) common to all families and listed in **??**; only the quantities flagged in each paragraph differ across families. Thus the $n = 80$ adherence and bridge experiments are catalogue-supervised at the target size; the held-out discovery rate is the stricter indicator of generalization beyond the supervised pool.

(i) Scratch GFlowNet at $n \in \{30, 50\}$ (architecture sweep). We train scratch GFlowNets at $n \in \{30, 50\}$ on five policy backbones (**??**): a feedforward MLP, a vanilla transformer encoder, and three configurations of the NT-Transformer with $(d_{\text{model}}, L) \in \{(128, 4), (256, 4), (256, 6)\}$, all with 8 primes and 16 Fourier frequency pairs. Each run minimizes the GFlowNet objective (**??**) for 10^4 updates with the canonical $80/20$ adherence split, the clipped-l^2 TB residual, $\log Z_\theta$ *fixed* to the ground-truth value $\log |\mathcal{M}_n|$ (i.e. initialised at the log-cardinality of the maximum-set family and *frozen* for the entire run, so that the trajectory-balance residual is driven only by the policy log-probabilities), and three random seeds.

(ii) Multi-n behavioural cloning evaluated at $n = 80$. We minimize the BC loss (**??**) of **??** on the multi-n expert pool $\bigcup_k \mathcal{A}_{n_k}$ with $n_k \in \{39, 40, 44, 49, 50, 55, 64, 65\}$, sampling one n_k uniformly per step. The backbone is the NT-Transformer swept over the five choices

$$(d_{\text{model}}, L) \in \{(128, 4), (128, 6), (256, 4), (256, 6), (256, 8)\},$$

with two learning rates $\{10^{-4}, 3 \times 10^{-4}\}$ and three random seeds per configuration. Optimization is AdamW with weight decay 10^{-2}. The trained policy is evaluated zero-shot at $n = 80$ against the full reference family $\mathcal{M}_{80}$; no GFlowNet term is used at any point.

(iii) Scratch GFlowNet at $n = 80$ (TB loss $\times$ $\log Z_\theta$ initialisation). We train a NT-Transformer backbone with $(d_{\text{model}}, L) \in \{(128, 4), (256, 4), (128, 6)\}$ at $n = 80$ under (**??**), loading $\mathcal{M}_{80}$ from a precomputed cache, with the canonical $80/20$ adherence split and three random seeds. $\log Z_\theta$ is trainable in every run. Two ablation axes are crossed: the TB residual form $\in$ {clipped-l^2, pseudo-Huber}, and the $\log Z_\theta$ initialisation $\in$ {default scalar, ground-truth $\log |\mathcal{M}_{80}|$ (GT)}. We label the four cells (iii.a) clipped-l^2 + default, (iii.b) clipped-l^2 + GT, (iii.c) pseudo-Huber + default, and (iii.d) pseudo-Huber + GT; (iii.a) inherits the canonical setup of (i)–(ii) extended to $n = 80$, while (iii.d) is the configuration we adopt as the default scratch baseline at $n = 80$ in subsequent paragraphs.

(iv) BC→bridge→GFlowNet at $n = 80$. We pretrain a NT-Transformer backbone with BC for $T_{\text{BC}} = 10^4$ steps on the multi-n pool of (ii), then fine-tune at $n^\star = 80$ for $T_{\text{FT}} = 10^4$ steps under the bridge loss (**??**) of **??**, with $\log Z_\theta$ initialized at $\log |\mathcal{M}_{80}|$ before fine-tuning. The adherence fraction at the target stays at the default 0.8. Four ablation axes are crossed: backbone $(d_{\text{model}}, L) \in \{(128, 4), (256, 4)\}$, TB residual form $\in$ {clipped-l^2, pseudo-Huber}, $\log Z_\theta$ trainability $\in$ {trainable, frozen}, and bridge regime — regime A ($T_{\text{bridge}} = 2000, \lambda_0 = 1$) versus regime B ($T_{\text{bridge}} = 800, \lambda_0 = 0.6$) — with three seeds.

(v) Curriculum transfer $30 \to 40 \to 50 \to 65 \to 80$. A single NT-Transformer policy is sequentially fine-tuned for 10^4 steps at each $n \in \{30, 40, 50, 65, 80\}$, with the previous stage's checkpoint as the initialization of the next stage and $\log Z_\theta$ reset to $\log |\mathcal{M}_n|$ at every transition (kept trainable). The objective at every stage is (**??**), and the adherence split at every stage is $80/20$. Three ablation axes are crossed: backbone $(d_{\text{model}}, L) \in \{(128, 4), (256, 4)\}$, TB residual form, and $\log Z_\theta$ trainability, with three seeds.

Metrics. We report four metrics consistent with **????**. The *coverage fraction* is the share of $\mathcal{M}_n$ that the policy has sampled at least once across the training run. The *relative entropy* of the empirical visit distribution over $\mathcal{M}_n$, normalized by $\log |\mathcal{M}_n|$, quantifies how uniformly the discovered solutions are revisited. The *held-out discovery rate* is the share of $\mathcal{H}_n$ that has been sampled at least once during training; this is our principal measure of generalization to maximum solutions never used by any training-time loss. Finally, the *adherence-vs-held-out log-p gap* measures the difference $\bar{\log} P(\mathcal{A}_n) - \bar{\log} P(\mathcal{H}_n)$ on ascending trajectories, with a small gap indicating that the model assigns comparable probabilities to in-pool and out-of-pool maximum solutions. All numbers are reported as the mean over three seeds with the seed-wise standard deviation in parentheses.

6. Results

We organize the empirical findings according to the five experimental families in **??**. This separates small-n architectural comparisons from the harder target-$n = 80$ setting, where we compare scratch GFlowNet training, zero-shot behavioural cloning, BC-initialized GFlowNet fine-tuning,

and curriculum transfer. **??** collects the full four-metric breakdown (coverage fraction, normalized entropy, held-out discovery rate, and the adherence-vs-held-out log-p gap) for every experimental family; the main-text paragraphs below summarize the headline numbers.

(i) Scratch GFlowNet architecture sweep at $n \in \{30, 50\}$. At $n = 30$ all backbones except NT-Transformer $(256, 6)$ saturate coverage and held-out discovery (1.000 ± 0.000); NT-Transformer $(256, 6)$ plateaus at $0.955\,(\pm 0.032)$. At $n = 50$ a clear ordering emerges: NT-Transformer $(128, 4)$ attains both the highest coverage (0.972 ± 0.003) and the highest held-out discovery rate (0.864 ± 0.015), 8.5 to 27 percentage points higher than the four other backbones (MLP 0.684, transformer 0.599, NT-Transformer $(256, 4)$ 0.779, NT-Transformer $(256, 6)$ 0.637). NT-Transformer $(256, 4)$ reaches comparable coverage but with the largest log-p gap (68.7 ± 3.7), indicating overfitting to $\mathcal{A}_{50}$; one of the three NT-Transformer $(256, 6)$ seeds collapses (coverage 0.25), driving the high seed-wise variance (std 0.34 on coverage). We adopt NT-Transformer $(128, 4)$ as the canonical small-n backbone for the $n = 80$ experiments that follow; the full four-metric breakdown is in **??**.

(ii) Multi-n behavioural cloning evaluated at $n = 80$. Across all five NT-Transformer configurations and both learning rates, supervised multi-n BC fails to transfer zero-shot to $n = 80$: coverage of $\mathcal{M}_{80}$ never exceeds 0.02% and only 4 of the 30 (architecture, lr, seed) cells produce any successful trajectory at all (**??** in **??**). The held-out discovery rate and adherence-vs-held-out log-p gap of (i) are undefined here, because BC never trains on $\mathcal{M}_{80}$ data, so the entire reference family at the target size is unseen and there is no $\mathcal{A}_{80}/\mathcal{H}_{80}$ split. We conclude that BC alone is insufficient at the target size; its value comes only from acting as a prior for subsequent GFlowNet fine-tuning, as analysed in (iv).

(iii) Scratch GFlowNet at $n = 80$. The scratch $n = 80$ experiments quantify the difficulty of on-policy GFlowNet training in the sparse maximum-solution regime. We first trained the NT-Transformer backbones under the historical clipped-l^2 TB residual with $\log Z_\theta$ initialised at the ground-truth value $\log |\mathcal{M}_{80}|$ — the configuration labelled (iii.b) (clipped-l^2 + GT init) in **??**, where "ground-truth" (abbreviated **GT** throughout) denotes the initialisation $\log Z_\theta \leftarrow \log |\mathcal{M}_{80}|$ at the known cardinality. At NT-Transformer $(256, 4)$ this already lifted coverage of $\mathcal{M}_{80}$ from essentially zero — the level reached by the zero-shot behavioural-cloning baseline of (ii) — to 0.235 ± 0.332. However, the seed-wise standard deviation exceeded the mean: two of three seeds collapsed entirely under the clipped-l^2 residual, leaving the result unreliable. We then switched to the pseudo-Huber residual under otherwise identical conditions — the configuration labelled (iii.d) (pseudo-Huber + GT init) — and observed a net improvement both in mean and in stability: coverage rose to 0.652 ± 0.043, and held-out discovery from 0.027 ± 0.038 to 0.119 ± 0.002 (**??** in **??**). The smaller backbones NT-Transformer $(128, 4)$ and $(128, 6)$ plateau at $\sim 10\%$ coverage under pseudo-Huber regardless of the $\log Z_\theta$ initialisation, indicating that capacity is the binding constraint at $n = 80$. Initialising $\log Z_\theta$ at GT versus the trainer's default scalar produces only a small additional improvement once the loss switch is applied.

(iv) BC→bridge→GFlowNet at $n = 80$. Combining the multi-n BC prior of (ii) with the GFlowNet objective of (iii) under the bridge schedule of **??** produces a large but catalogue-supervised improvement over both component baselines: total coverage of $\mathcal{M}_{80}$ rises to 0.82–0.83 across all 16 ablation cells (versus 0.65 at the best scratch cell of (iii) and essentially zero for zero-shot BC alone), while held-out discovery reaches 0.18–0.21 at NT-Transformer $(128, 4)$ and 0.12–0.14 at NT-Transformer $(256, 4)$ (**??** in **??**). Because this setting uses 80% of the $n = 80$ catalogue through the adherence/bridge losses, the held-out rate rather than total coverage is the stricter measure of discovery beyond supervised target-size solutions. The three secondary axes — TB residual form, $\log Z_\theta$ trainability, and bridge regime — collapse to within seed noise once the BC prior is in place: within each architecture the four (loss $\times$ $\log Z_\theta$ $\times$ bridge) cells differ by less than 0.013 in coverage with seed-wise standard deviations of 0.001–0.012. Architecture remains the dominant remaining axis: the wider NT-Transformer $(256, 4)$ attains higher relative entropy on the discovered subset (≈ 0.955 vs ≈ 0.88, closer to uniform) but pays for it with a held-out discovery rate roughly 6 percentage points lower and a log-p gap of $\approx$ 100–107 nats — over three times the $\approx$ 27–32 nat gap of NT-Transformer $(128, 4)$. The wider backbone samples the adherence subset more uniformly but generalises worse to held-out maximum solutions, consistent with the overfitting tendency already observed at (iii) under scratch training.

(v) Curriculum transfer $30 \to 40 \to 50 \to 65 \to 80$. We report two configurations of the NT-Transformer at the final $n = 80$ stage, both under the pseudo-Huber TB residual (**??** in **??**). The smaller backbone $(128, 4)$ with frozen $\log Z_\theta$ reaches coverage 0.224 ± 0.083, an order of magnitude below its BC-bridge counterpart of (iv) and below the best scratch cell of (iii). The larger backbone $(256, 4)$ with trainable $\log Z_\theta$ reaches coverage 0.771 ± 0.025 — markedly above its best scratch cell (0.652, (iii.d)) and approaching BC→bridge (0.82–0.83, (iv)). Held-out discovery at the larger backbone under the curriculum (0.132 ± 0.003) matches BC→bridge at the same architecture (0.12–0.14), indicating that the multi-n curriculum supplies a comparable transfer signal to BC at this size.

The five families together answer the two questions of **??**. *First,* scratch GFlowNet training at $n = 80$ is feasible but fragile: under the historical clipped-l^2 TB residual it collapses on most seeds, and under the robust pseudo-Huber it tops out at coverage ≈ 0.65 with held-out discovery below 0.12. *Second,* transfer signals provide a clear and reliable lift: BC→bridge fine-tuning raises total coverage of $\mathcal{M}_{80}$ to 0.82–0.83 across the full 16-cell ablation grid and pushes held-out discovery to 0.18–0.21 at the smaller backbone, with the loss/$\log Z_\theta$/bridge axes collapsing to seed noise once the BC prior is in place; the curriculum, on the partial evidence available, also surpasses scratch at the larger backbone but does not reach BC-bridge. Across both transfer pipelines architecture remains the principal axis: NT-Transformer $(256, 4)$ samples the adherence subset more uniformly but generalises worse to held-out solutions than NT-Transformer $(128, 4)$, a consistent uniformity-vs-generalization trade-off.

7. Conclusion and future work

We have presented a controlled first step towards using GFlowNets to construct families of combinatorial objects, instantiated here on the problem of sampling maximum Sidon sets. The three ingredients we study — a strong inductive bias, an adherence loss, and behavioural cloning — were chosen because they are compatible with future settings where only sparse verified reference families are available. Across our experiments the dominant lesson is that *inductive bias is fundamental*: the NT-Transformer backbone, which exposes per-element modular and Fourier structure together with a sum-conflict count encoding the Sidon constraint, consistently dominates the MLP and vanilla-transformer baselines under matched training conditions. The two supervision channels we test exploit partial verified information: the *adherence loss* leverages whatever partial catalogue of maximum sets happens to be known at the target n^*, and *behavioural cloning* transfers from enumerated families at sizes $n < n^*$ to the unseen target n^*. Both prove to be effective supplements to on-policy training at $n = 80$.

Limitations. Our evaluation should be interpreted as a preliminary controlled study. All quantitative settings remain enumerable or catalogue-based, the strongest $n = 80$ model uses 80% of the target-size catalogue during training, and held-out discovery remains much lower than total catalogue coverage at this stage.

Future work. Our next steps focus on scaling this methodology to the substantially harder $n^* \in \{200, 500\}$ regimes, where $\mathcal{M}_{n^*}$ can no longer be enumerated and the on-policy GFlowNet signal is correspondingly sparser. A second evaluation priority is to compare against optimized randomized backtracking and other exact-search baselines under matched rollout or node-expansion budgets. Two changes to the training objective are motivated directly by the observations of this paper. First, we will remove the dead-end penalty and replace the constant terminal reward used here with a *size-scaled reward* that grants non-zero credit to Sidon sets of size below $M(n^*)$ in proportion to their cardinality: our experiments suggest that the absence of a continuous learning signal across submaximal terminals is detrimental to training. Second, we will integrate additional behavioural cloning and curriculum learning recipes, alongside data-enhancement techniques such as PatternBoost (**?**), into the same masked-sequential framework.

References

Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatain, M., Novikov, A., Ruiz, F. J. R., Schrittwieser, J., Swirszcz, G., Silver, D., Hassabis, D., and Kohli, P. 2022. Discovering Faster Matrix Multiplication Algorithms with Reinforcement Learning. *Nature*, 610(7930):47–53.

Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., Ruiz, F. J. R., Ellenberg, J. S., Wang, P., Fawzi, O., Kohli, P., and Fawzi, A. 2024. Mathematical Discoveries from Program Search with Large Language Models. *Nature*, 625(7995):468–475.

O'Bryant, K. 2004. A Complete Annotated Bibliography of Work Related to Sidon Sequences. *Electronic Journal of Combinatorics*, DS11.

Balogh, J., Füredi, Z., and Roy, S. 2023. An Upper Bound on the Size of Sidon Sets. *American Mathematical Monthly*, 130(5):437–445.

Etzion, T. 2011. Sidon Sequences and Doubly Periodic Two-Dimensional Synchronization Patterns. In *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 144–148.

Bengio, E., Jain, M., Korablyov, M., Precup, D., and Bengio, Y. 2021. Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation. In *Advances in Neural Information Processing Systems*, pp. 27381–27394.

Bengio, Y., Lahlou, S., Deleu, T., Hu, E. J., Tiwari, M., and Bengio, E. 2023. GFlowNet Foundations. *Journal of Machine Learning Research*, 24(210):1–55.

Malkin, N., Jain, M., Bengio, E., Sun, C., and Bengio, Y. 2022. Trajectory Balance: Improved Credit Assignment in GFlowNets. In *Advances in Neural Information Processing Systems*, pp. 5955–5967.

Madan, K., Rector-Brooks, J., Korablyov, M., Bengio, E., Jain, M., Nica, A. C., Bosc, T., Bengio, Y., and Malkin, N. 2023. Learning GFlowNets from Partial Episodes for Improved Convergence and Stability. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*.

Zhang, D., Dai, H., Malkin, N., Courville, A. C., Bengio, Y., and Pan, L. 2023. Let the Flows Tell: Solving Graph Combinatorial Optimization Problems with GFlowNets. In *Advances in Neural Information Processing Systems*.

Charton, F., Ellenberg, J. S., Wagner, A. Z., and Williamson, G. 2024. PatternBoost: Constructions in Mathematics with a Little Help from AI. *arXiv preprint arXiv:2411.00566*.

A. Detailed training configuration

?? reports the training hyperparameters that are common to every experimental family of **??**; quantities that vary per family are stated in the corresponding paragraph and are not repeated here.

Table 1. Shared training configuration. Quantities that vary per family of experiments are stated in the corresponding paragraph of **??**; everything else is held fixed at the values below.

Updates per training run	10^4
Rollout batch size B	256
Optimizer	Adam (BC and bridge BC: AdamW, wd $= 10^{-2}$)
Learning rate	10^{-4} (warmup 200 + cosine to 10^{-6})
β_{dead}	2.0
β_{ent}	0.01
Adherence fraction (default)	0.8
Adherence anneal T_{anneal}	T_{train} (full run)
Adherence mini-batch	256
Sampling temperature τ	$2.0 \to 1.0$ (linear)
Mixture floor ε	$0.3 \to 0.05$ (linear)
Pseudo-Huber TB δ	20
Split seed	42
Random seeds per setting	3
Gradient clip	1.0

B. Reproducibility details for catalogues and splits

The experiments rely on exact reference catalogues for the reported enumerable sizes. To make the catalogue-supervised protocol explicit, each run is determined by the ambient size n, the reference family $\mathcal{R}_n$, the split seed in **??**, and the experimental family in **??**. For the target-size experiments at $n = 80$, the canonical split uses 80% of $\mathcal{M}_{80}$ as the adherence pool $\mathcal{A}_{80}$ and reserves the remaining 20% as $\mathcal{H}_{80}$, which is never used by the adherence or BC losses.

Catalogue construction. The catalogues are generated by exact ordered enumeration: the search builds increasing Sidon prefixes, maintains the realized pairwise-sum set $\Sigma(s)$, branches only on feasible insertions that preserve the Sidon constraint, and records a terminal set only when it has cardinality $M(n)$. This is the same feasibility rule used by the GFlowNet mask, but run exhaustively rather than stochastically. Duplicate trajectories are avoided by the increasing-order convention.

Splits and budgets. The adherence/held-out partition is produced once per size using split seed 42 and then held fixed across all methods, so differences between methods are not due to different train/test partitions. Each GFlowNet run uses 10^4 updates with rollout batch size 256; hence coverage curves and final coverage fractions aggregate discoveries over approximately 2.56×10^6 training rollouts per seed. BC-only runs are evaluated zero-shot at $n = 80$ and never use $\mathcal{A}_{80}$ or $\mathcal{H}_{80}$ during training.

C. Per-experiment metric tables

This appendix collects the full four-metric comparison for every experimental family of **??**: coverage fraction (↑), normalized entropy of the empirical visit distribution (↑), held-out discovery rate (↑), and the adherence-vs-held-out log-p gap (closer to 0 is better). All numbers are reported at the end of training as the mean over three random seeds with the seed-wise standard deviation in parentheses. All runs share the schedule of **??**, the $80/20$ adherence split, and the reproducibility protocol in **??**. We begin with the architecture sweep of family (i); tables for families (ii)–(v) follow the same template.

Table 2. Family (i) — architecture sweep at $n = 30$. Mean $\pm$ seed-wise std over three seeds.

Architecture	Cov. frac. ↑	Rel. entropy ↑	H.O. disc. rate ↑	log-p gap ($\rightarrow 0$)
mlp	1.000 ± 0.000	0.967 ± 0.001	1.000 ± 0.000	1.92 ± 0.12
transformer	1.000 ± 0.001	0.984 ± 0.001	0.998 ± 0.003	6.07 ± 0.34
nt-T $(128, 4)$	1.000 ± 0.000	0.974 ± 0.000	1.000 ± 0.000	4.28 ± 0.96
nt-T $(256, 4)$	1.000 ± 0.000	0.978 ± 0.003	1.000 ± 0.000	11.76 ± 8.40
nt-T $(256, 6)$	0.955 ± 0.032	0.966 ± 0.004	0.961 ± 0.028	2.27 ± 3.28

Table 3. Family (i) — architecture sweep at $n = 50$. Mean $\pm$ seed-wise std over three seeds. Note the collapsed seed for nt-T $(256, 6)$ that drives the high std on coverage and held-out discovery rate.

Architecture	Cov. frac. ↑	Rel. entropy ↑	H.O. disc. rate ↑	log-p gap ($\rightarrow 0$)
mlp	0.937 ± 0.003	0.917 ± 0.001	0.684 ± 0.015	9.07 ± 0.61
transformer	0.919 ± 0.006	0.969 ± 0.001	0.599 ± 0.027	21.10 ± 1.64
nt-T $(128, 4)$	0.972 ± 0.003	0.877 ± 0.009	0.864 ± 0.015	14.50 ± 1.34
nt-T $(256, 4)$	0.954 ± 0.001	0.956 ± 0.002	0.779 ± 0.007	68.68 ± 3.66
nt-T $(256, 6)$	0.726 ± 0.335	0.894 ± 0.086	0.637 ± 0.251	44.84 ± 31.75

Table 4. Family (ii) — multi-n BC pretrained on $\bigcup_k \mathcal{A}_{n_k}$ with $n_k \in \{39, 40, 44, 49, 50, 55, 64, 65\}$ and evaluated zero-shot at $n = 80$. 5 architectures $\times$ 2 learning rates $\times$ 3 seeds (mean $\pm$ seed-wise std). The held-out discovery rate and adherence-vs-held-out log-p gap reported for family (i) are undefined here because BC never trains on $\mathcal{M}_{80}$ data: the full reference family at the target size is unseen and there is no $\mathcal{A}_{80}/\mathcal{H}_{80}$ split, so coverage on $\mathcal{M}_{80}$ already serves as the generalization measure. Coverage and success rate are essentially 0 across the grid.

Architecture	lr	Cov. frac. ↑	Rel. entropy ↑	Success rate ↑
nt-T $(128, 4)$	1×10^{-4}	0.0002 ± 0.0002	0.000 ± 0.000	0.0003 ± 0.0005
nt-T $(128, 4)$	3×10^{-4}	0.0002 ± 0.0002	0.000 ± 0.000	0.0003 ± 0.0005
nt-T $(128, 6)$	1×10^{-4}	0.0000 ± 0.0000	0.000 ± 0.000	0.0000 ± 0.0000
nt-T $(128, 6)$	3×10^{-4}	0.0002 ± 0.0002	0.000 ± 0.000	0.0003 ± 0.0005
nt-T $(256, 4)$	1×10^{-4}	0.0000 ± 0.0000	0.000 ± 0.000	0.0000 ± 0.0000
nt-T $(256, 4)$	3×10^{-4}	0.0000 ± 0.0000	0.000 ± 0.000	0.0000 ± 0.0000
nt-T $(256, 6)$	1×10^{-4}	0.0000 ± 0.0000	0.000 ± 0.000	0.0000 ± 0.0000
nt-T $(256, 6)$	3×10^{-4}	0.0000 ± 0.0000	0.000 ± 0.000	0.0000 ± 0.0000
nt-T $(256, 8)$	1×10^{-4}	0.0000 ± 0.0000	0.000 ± 0.000	0.0000 ± 0.0000
nt-T $(256, 8)$	3×10^{-4}	0.0002 ± 0.0002	0.000 ± 0.000	0.0003 ± 0.0005

Table 5. Family (iii) — scratch GFlowNet at $n = 80$. Each row averages three seeds (mean $\pm$ seed-wise std) on the canonical $80/20$ adherence split. The top half uses the historical clipped-l^2 TB residual; the bottom half uses the pseudo-Huber residual. $\log Z_\theta$ is trainable in every row. Rows tagged *GT* (ground-truth) initialise $\log Z_\theta \leftarrow \log |\mathcal{M}_{80}|$ before training; the other rows initialise $\log Z_\theta$ at the trainer's default scalar. Note the large seed-wise std on every clipped-l^2 row, driven by collapsed seeds that pseudo-Huber eliminates.

TB loss	$\log Z_\theta$ init	Arch	Cov. frac. ↑	Rel. entropy ↑	H.O. disc. ↑
clipped-l^2	default (iii.a)	nt-T $(128, 4)$	0.036 ± 0.051	0.233 ± 0.330	0.024 ± 0.035
clipped-l^2	default (iii.a)	nt-T $(256, 4)$	0.263 ± 0.372	0.292 ± 0.414	0.034 ± 0.048
clipped-l^2	default (iii.a)	nt-T $(128, 6)$	0.002 ± 0.003	0.105 ± 0.149	0.001 ± 0.001
clipped-l^2	GT (iii.b)	nt-T $(128, 4)$	0.040 ± 0.057	0.239 ± 0.338	0.025 ± 0.036
clipped-l^2	GT (iii.b)	nt-T $(256, 4)$	0.235 ± 0.332	0.281 ± 0.397	0.027 ± 0.038
clipped-l^2	GT (iii.b)	nt-T $(128, 6)$	0.037 ± 0.052	0.235 ± 0.333	0.030 ± 0.043
pseudo-Huber	default (iii.c)	nt-T $(128, 4)$	0.100 ± 0.012	0.690 ± 0.016	0.086 ± 0.010
pseudo-Huber	default (iii.c)	nt-T $(256, 4)$	0.677 ± 0.039	0.793 ± 0.022	0.097 ± 0.008
pseudo-Huber	default (iii.c)	nt-T $(128, 6)$	0.114 ± 0.008	0.707 ± 0.009	0.105 ± 0.007
pseudo-Huber	GT (iii.d)	nt-T $(128, 4)$	0.101 ± 0.005	0.692 ± 0.007	0.086 ± 0.011
pseudo-Huber	GT (iii.d)	nt-T $(256, 4)$	0.652 ± 0.043	0.794 ± 0.018	0.119 ± 0.002
pseudo-Huber	GT (iii.d)	nt-T $(128, 6)$	0.108 ± 0.011	0.700 ± 0.014	0.085 ± 0.016

Table 6. Family (iv) — BC→bridge→GFlowNet at $n = 80$. Each row averages three seeds (mean $\pm$ seed-wise std) on the canonical $80/20$ adherence split. Regime A is the long-linear bridge (T_{bridge}=2000, λ_0=1.0); regime B is the short-soft bridge (T_{bridge}=800, λ_0=0.6). Coverage of $\mathcal{M}_{80}$ is uniformly high across all 16 cells; the three secondary axes (TB loss, $\log Z_\theta$ trainability, bridge regime) collapse to seed-noise, while architecture remains a live axis with a clear coverage-vs-generalization trade-off.

Arch	Bridge	TB loss	$\log Z_\theta$	Cov. frac. ↑	Rel. ent. ↑	H.O. disc. ↑	log-p gap (→ 0)
nt-T $(128, 4)$	A	clipped-l^2	frozen	0.831 ± 0.006	0.881 ± 0.008	0.181 ± 0.016	27.5 ± 0.5
nt-T $(128, 4)$	A	clipped-l^2	trainable	0.831 ± 0.003	0.894 ± 0.005	0.181 ± 0.005	27.2 ± 0.6
nt-T $(128, 4)$	A	pseudo-Huber	frozen	0.818 ± 0.012	0.857 ± 0.010	0.198 ± 0.019	31.4 ± 1.0
nt-T $(128, 4)$	A	pseudo-Huber	trainable	0.824 ± 0.002	0.869 ± 0.008	0.198 ± 0.014	32.2 ± 0.9
nt-T $(128, 4)$	B	clipped-l^2	frozen	0.831 ± 0.006	0.882 ± 0.009	0.190 ± 0.018	27.3 ± 0.7
nt-T $(128, 4)$	B	clipped-l^2	trainable	0.832 ± 0.003	0.895 ± 0.006	0.180 ± 0.018	27.5 ± 0.8
nt-T $(128, 4)$	B	pseudo-Huber	frozen	0.821 ± 0.007	0.855 ± 0.009	0.194 ± 0.006	31.4 ± 0.4
nt-T $(128, 4)$	B	pseudo-Huber	trainable	0.825 ± 0.006	0.869 ± 0.009	0.208 ± 0.009	31.1 ± 0.8
nt-T $(256, 4)$	A	clipped-l^2	frozen	0.824 ± 0.003	0.957 ± 0.001	0.127 ± 0.012	101.8 ± 1.0
nt-T $(256, 4)$	A	clipped-l^2	trainable	0.823 ± 0.002	0.958 ± 0.001	0.122 ± 0.006	99.2 ± 2.3
nt-T $(256, 4)$	A	pseudo-Huber	frozen	0.828 ± 0.004	0.953 ± 0.001	0.140 ± 0.019	106.5 ± 2.0
nt-T $(256, 4)$	A	pseudo-Huber	trainable	0.826 ± 0.003	0.955 ± 0.001	0.133 ± 0.017	107.2 ± 1.8
nt-T $(256, 4)$	B	clipped-l^2	frozen	0.824 ± 0.002	0.957 ± 0.000	0.127 ± 0.008	102.1 ± 3.1
nt-T $(256, 4)$	B	clipped-l^2	trainable	0.825 ± 0.002	0.958 ± 0.000	0.128 ± 0.010	101.0 ± 2.0
nt-T $(256, 4)$	B	pseudo-Huber	frozen	0.824 ± 0.001	0.953 ± 0.001	0.119 ± 0.007	107.7 ± 0.3
nt-T $(256, 4)$	B	pseudo-Huber	trainable	0.826 ± 0.002	0.955 ± 0.001	0.129 ± 0.012	105.8 ± 1.7

Table 7. Family (v) — curriculum transfer $30 \to 40 \to 50 \to 65 \to 80$, final-stage ($n = 80$) metrics for two configurations of the NT-Transformer under the pseudo-Huber TB residual: $(128, 4)$ with frozen $\log Z_\theta$ and $(256, 4)$ with trainable $\log Z_\theta$, two seeds each. Mean $\pm$ seed-wise std on the canonical $80/20$ adherence split.

Arch	TB loss	$\log Z_\theta$	Seeds	Cov. frac. ↑	Rel. ent. ↑	H.O. disc. ↑	log-p gap (→ 0)
nt-T $(128, 4)$	pseudo-Huber	frozen	2	0.224 ± 0.083	0.777 ± 0.043	0.134 ± 0.010	7.77 ± 2.37
nt-T $(256, 4)$	pseudo-Huber	trainable	2	0.771 ± 0.025	0.853 ± 0.012	0.132 ± 0.003	39.46 ± 6.57

D. Sizes of the maximum Sidon-set family $\mathcal{M}_N$

?? reports the cardinality $|\mathcal{M}_N|$ and the maximum Sidon-set size $M(N)$ for ambient sizes $N \in [100, 151]$. These values can be used to initialize the partition function, for example by setting $\log Z_\theta \leftarrow \log |\mathcal{M}_N|$, before further fine-tuning $\log Z_\theta$ jointly with the policy (see **??**).

Table 8. Sizes of the maximum Sidon-set family $\mathcal{M}_N$ for several ambient sizes N. Here $k^* = M(N)$ is the maximum Sidon-set cardinality and `num_sets` is $|\mathcal{M}_N|$.

N	k^*	`num_sets`	N	k^*	`num_sets`
100	12	12,096	128	14	2
105	12	251,254	129	14	6
106	12	432,940	130	14	16
107	13	2	131	14	28
108	13	4	132	14	52
109	13	6	133	14	84
110	13	14	134	14	136
111	13	28	135	14	240
112	13	60	136	14	434
113	13	134	137	14	820
114	13	264	138	14	1,582
115	13	538	139	14	2,962
116	13	1,106	140	14	5,644
117	13	2,228	141	14	10,524
118	13	4,478	142	14	19,330
119	13	8,706	143	14	34,790
120	13	16,734	144	14	62,138
121	13	30,950	145	14	108,630
122	13	56,692	146	14	188,858
123	13	100,822	147	14	321,182
124	13	177,448	148	14	540,724
125	13	305,430	149	14	894,356
126	13	521,012	150	14	1,468,600
127	13	869,342	151	14	2,372,890

E. Training curves at $n = 80$

This appendix shows the seed-mean coverage-fraction curves over training steps for the $n = 80$ pipelines discussed in families (iii) and (iv) of **??**. Each curve is the mean across the three random seeds; the shaded band is the seed-wise standard deviation. All runs use the canonical $80/20$ adherence split.

Scratch GFlowNet under clipped-l^2 TB residual. **??** shows coverage of $\mathcal{M}_{80}$ over 10^4 training steps for the three NT-Transformer backbones $(d_{\text{model}}, L) \in \{(128, 4), (256, 4), (128, 6)\}$ under the historical clipped-l^2 TB residual, including both the default $\log Z_\theta$ initialisation and the ground-truth init (cells (iii.a) and (iii.b)). Note the very wide shaded bands, driven by collapsed seeds on most cells.

Scratch GFlowNet under pseudo-Huber TB residual. **??** shows the same backbones retrained under pseudo-Huber TB (cells (iii.c) and (iii.d)), together with two exploration sweeps (*lvl1*: $\tau_{\text{end}} = 1.15$, $\varepsilon_{\text{end}} = 0.08$; *lvl2*: $\tau_{\text{end}} = 1.30$, $\varepsilon_{\text{end}} = 0.12$) on NT-Transformer $(128, 4)$. The bands are markedly tighter than under clipped-l^2, confirming the seed-collapse fix.

BC→bridge→GFlowNet. **????????** show coverage of $\mathcal{M}_{80}$ across the BC pretrain (steps 0–10^4), the bridge phase (first portion of the fine-tuning window), and the pure-GFlowNet phase. The four panels span the two backbones $(d_{\text{model}}, L) \in \{(128, 4), (256, 4)\}$ and the two bridge regimes A (long-linear, $T_{\text{bridge}} = 2000$, $\lambda_0 = 1$) and B (short-soft, $T_{\text{bridge}} = 800$, $\lambda_0 = 0.6$). All four cells reach coverage ≈ 0.82–0.83, with seed-wise bands tighter than 0.013 on every cell.

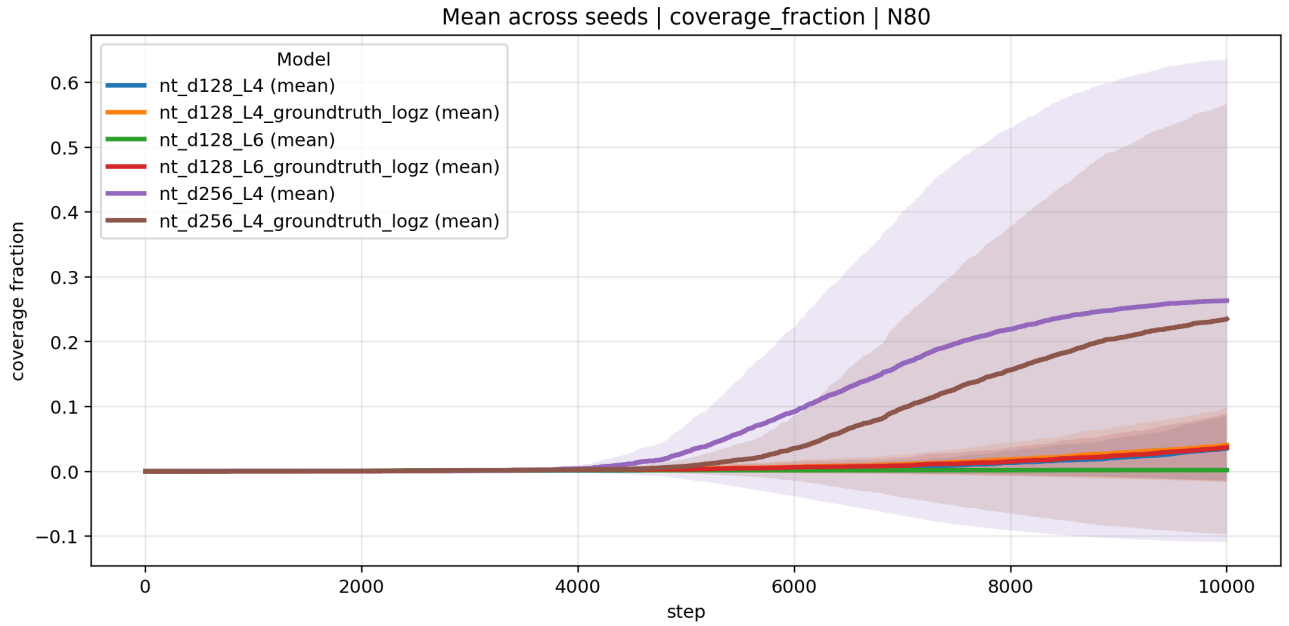

Figure 1. Family (iii) scratch GFlowNet at $n = 80$ under the clipped-l^2 TB residual. Coverage of $\mathcal{M}_{80}$ vs. training step, mean across three seeds with seed-wise standard-deviation band.

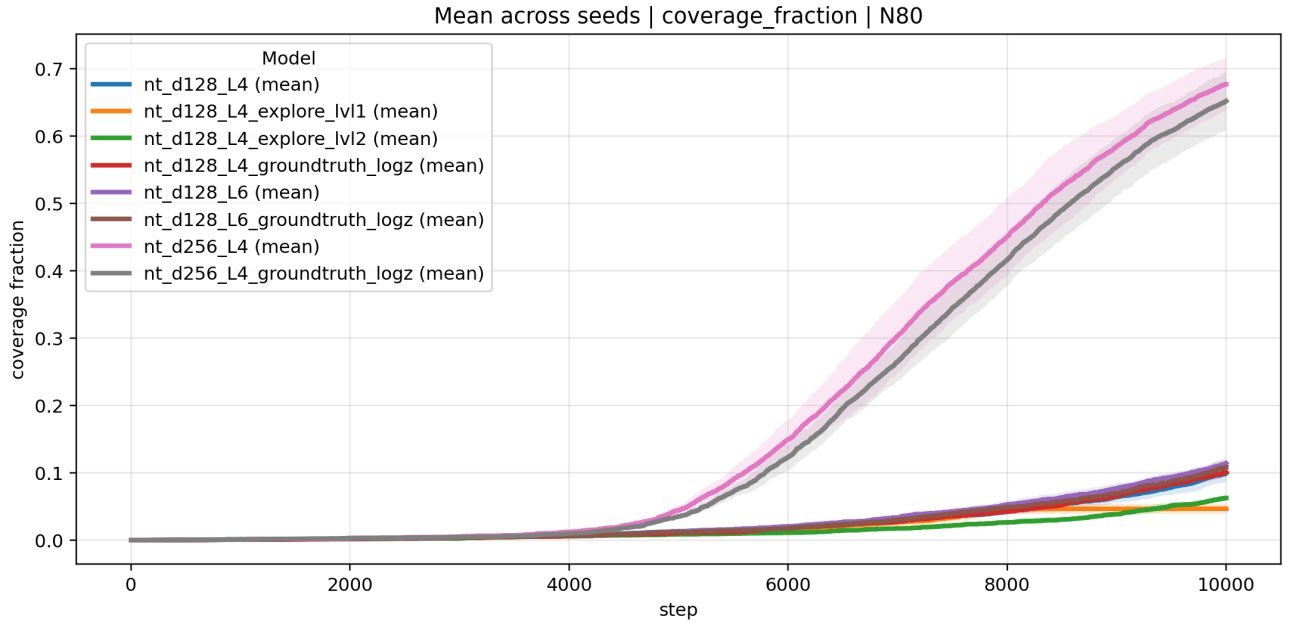

Figure 2. Family (iii) scratch GFlowNet at $n = 80$ under the pseudo-Huber TB residual, with two exploration-sweep variants on NT-Transformer $(128, 4)$. Coverage of $\mathcal{M}_{80}$ vs. training step, mean across three seeds with seed-wise standard-deviation band.

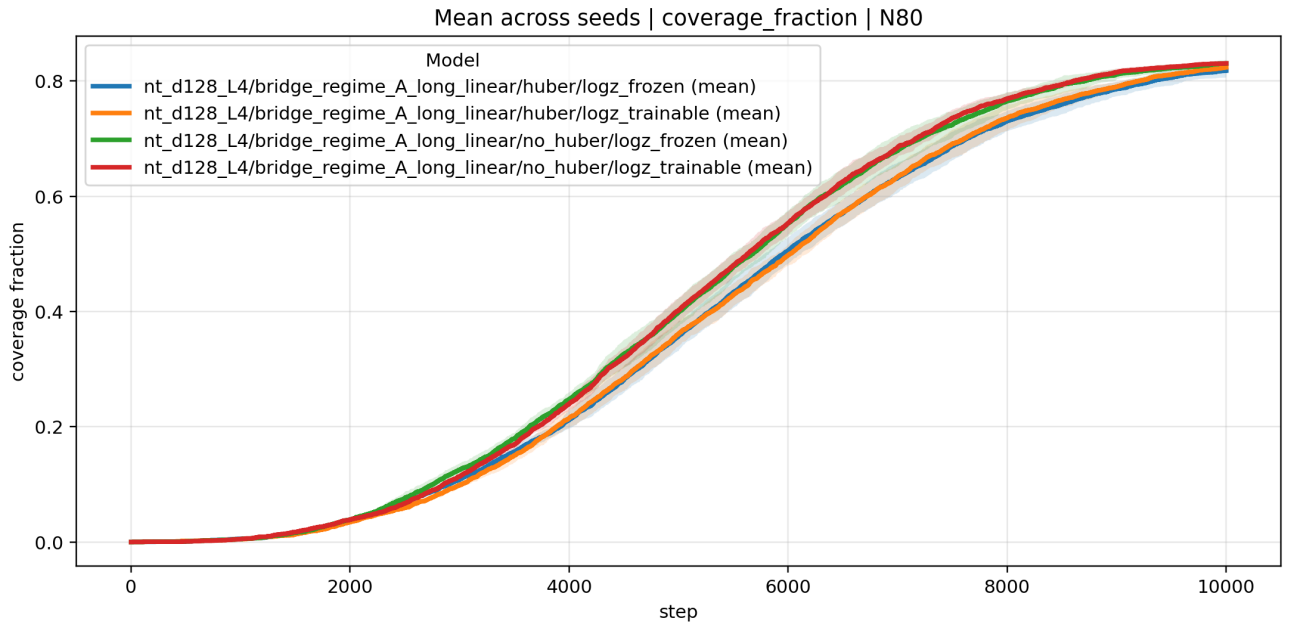

Figure 3. Family (iv): BC→bridge→GFlowNet at $n = 80$ for NT-Transformer $(128, 4)$ under regime A (long-linear bridge, $T_{\text{bridge}} = 2000$, $\lambda_0 = 1$). Pseudo-Huber TB residual, frozen $\log Z_\theta$ initialised at $\log |\mathcal{M}_{80}|$.

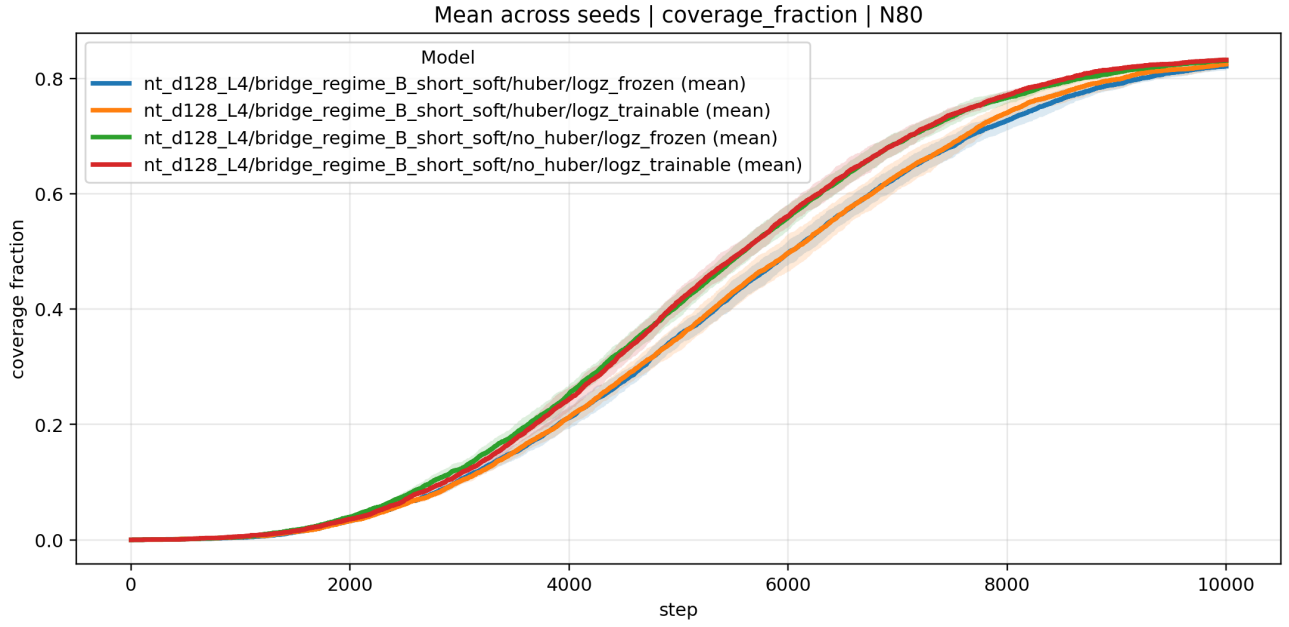

Figure 4. Family (iv): BC→bridge→GFlowNet at $n = 80$ for NT-Transformer $(128, 4)$ under regime B (short-soft bridge, $T_{\text{bridge}} = 800$, $\lambda_0 = 0.6$). Pseudo-Huber TB residual, frozen $\log Z_\theta$ initialised at $\log |\mathcal{M}_{80}|$.

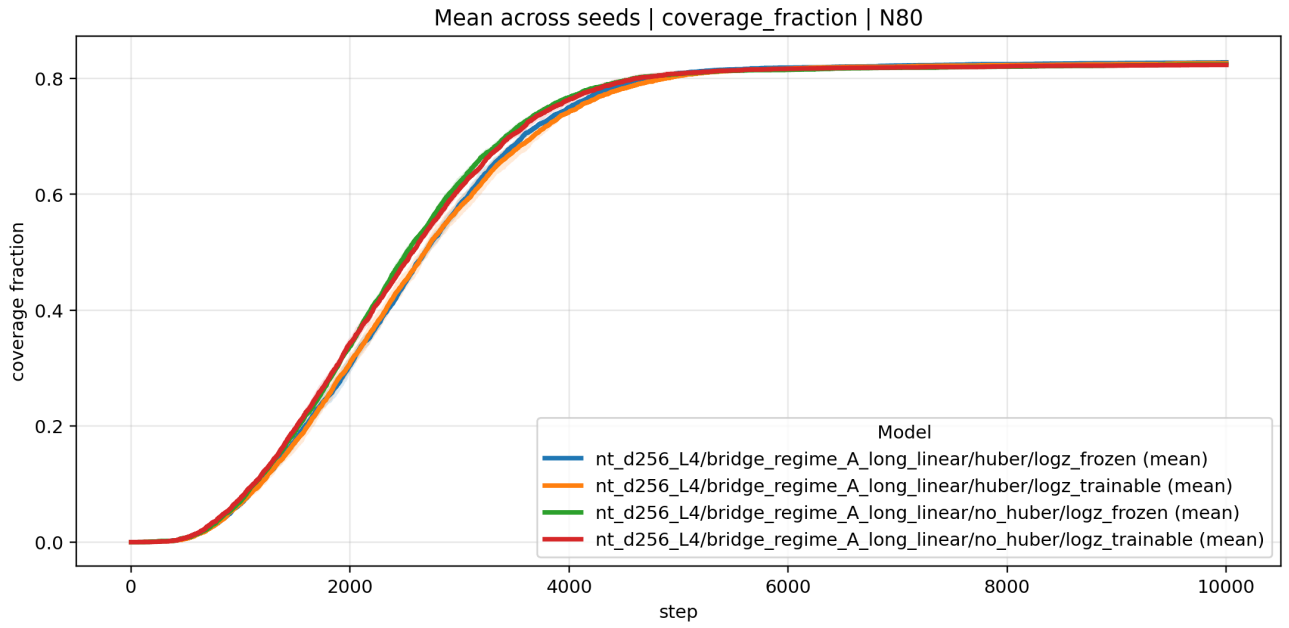

Figure 5. Family (iv): BC→bridge→GFlowNet at $n = 80$ for NT-Transformer $(256, 4)$ under regime A (long-linear bridge, $T_{\text{bridge}} = 2000$, $\lambda_0 = 1$). Pseudo-Huber TB residual, frozen $\log Z_\theta$ initialised at $\log |\mathcal{M}_{80}|$.

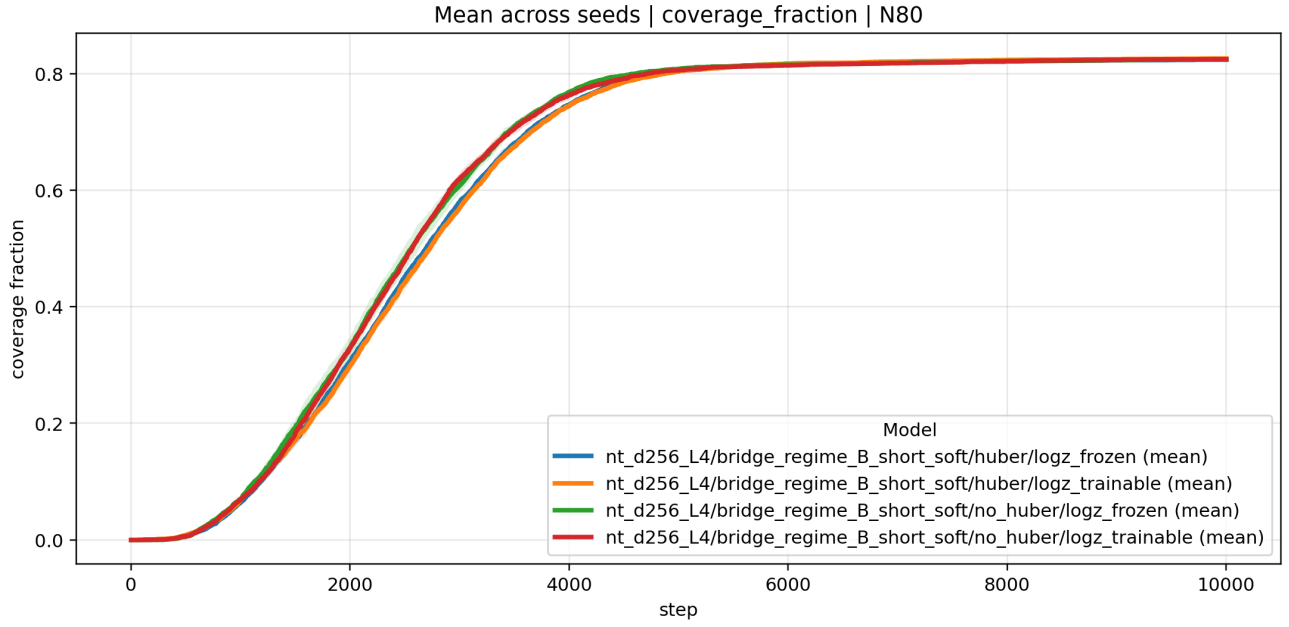

Figure 6. Family (iv): BC→bridge→GFlowNet at $n = 80$ for NT-Transformer $(256, 4)$ under regime B (short-soft bridge, $T_{\text{bridge}} = 800$, $\lambda_0 = 0.6$). Pseudo-Huber TB residual, frozen $\log Z_\theta$ initialised at $\log |\mathcal{M}_{80}|$.