arXiv: 2605.18360

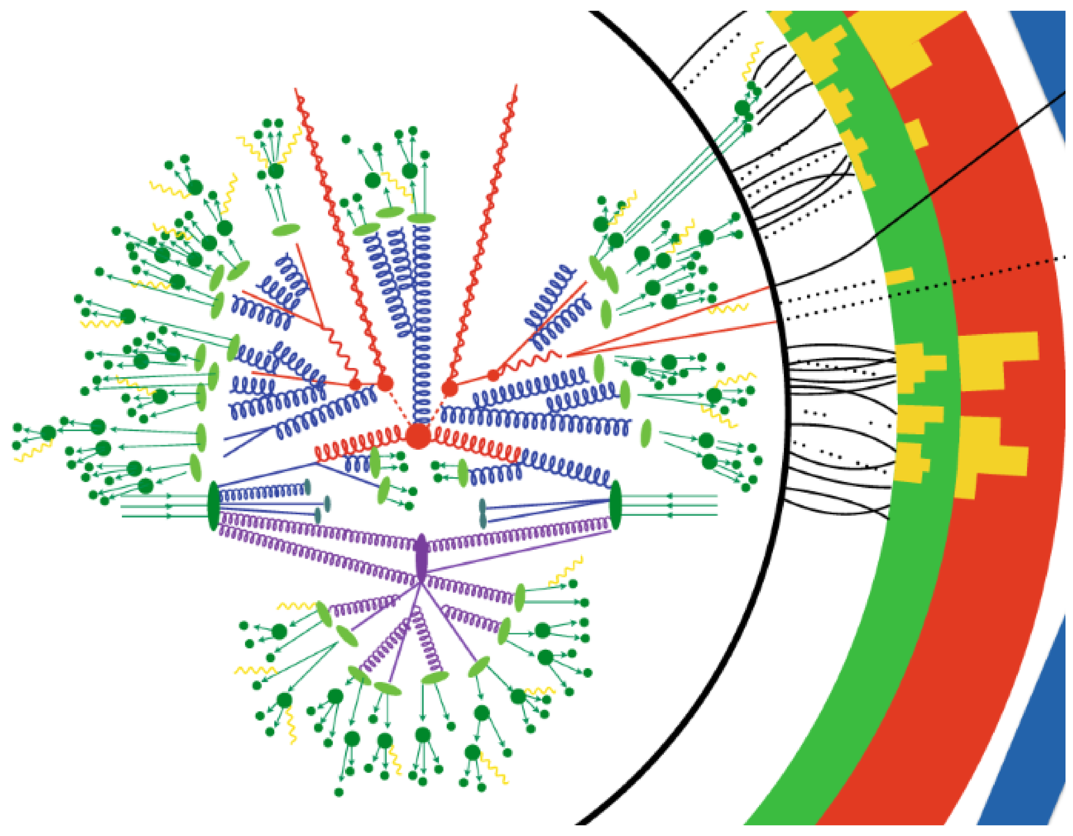
Nested-GPT for variable-multiplicity parton showers: A case study in the resummation of non-global logarithms

Wanchen Li, Ding Yu Shao, Hao-Zhe Shi, Yu-Xuan Sun

Department of Physics and Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China;
Key Laboratory of Nuclear Physics and Ion-beam Application (MOE);
Shanghai Research Center for Theoretical Nuclear Physics;
Center for High Energy Physics, Peking University; SCNT, IMP-CAS.

Parton Showers

Parton showers are the QCD radiation engine of Monte Carlo event generators. They turn a hard scattering into an exclusive final state by simulating ordered soft and collinear emissions.



- **Precision SM:** jet vetoes, event shapes, and substructure depend on shower fidelity.
- **New-physics baseline:** QCD radiation sets the background and uncertainty floor.
- **Computational pressure:** NLL/NNLL showers require expensive high-statistics simulations.

Aim: learn ordered shower histories once, then generate high-statistics samples faster for precision analysis.

Non-Global Logarithms

Non-global logarithms (NGLs) arise when radiation is vetoed only in part of phase space. Emissions outside the veto region can radiate back into it, so the survival probability is driven by correlated, nonlinear branching.

$$Q\frac{\partial G_{ab}}{\partial Q} = \int \frac{d^2\Omega_k}{4\pi}\,\bar{\alpha}_s W_{ab}(k)\,\left[u(k)G_{ak}G_{kb} - G_{ab}\right]$$

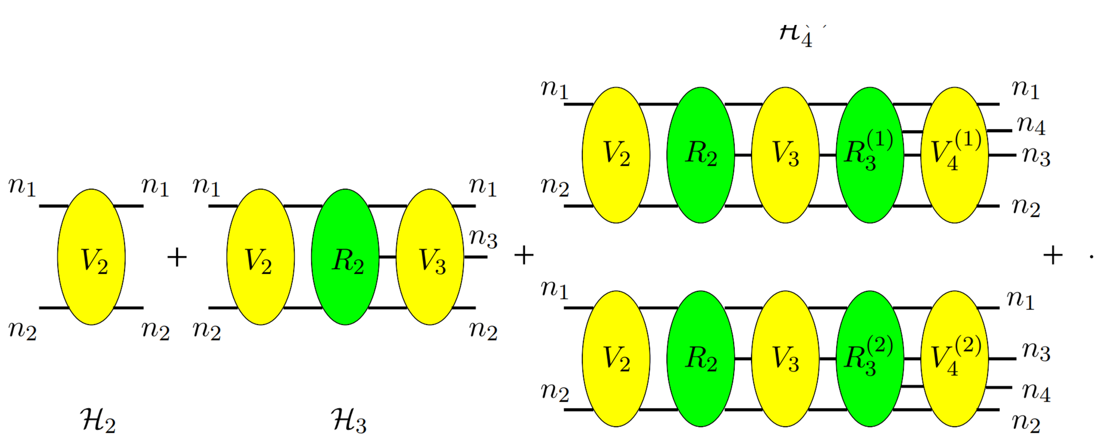


Figure. virtual evolution and real emissions recursively build the ordered shower history.

- Real emissions split one dipole into two daughter dipoles; virtual evolution supplies the Sudakov no-emission probability.
- If an emission enters the veto gap, the event terminates; otherwise it is inserted into the ordered dipole list.
- NGLs are therefore a compact but demanding benchmark for variable-length, time-ordered generation.

Outside gap	insert the emission into the dipole chain and continue the shower
Inside gap	stop the event and record a veto failure
No emission	accumulate Sudakov survival through virtual evolution

The challenge is not just particle kinematics, but causal branching plus a learned stopping rule.

Generative AI

- Train on reference shower histories generated by the LL large-N_c Monte Carlo solution.
- Keep the physically relevant sequence information: shower time, emission kinematics, and event termination.
- Compare a fixed-multiplicity flow-matching baseline with an autoregressive model that generates until it chooses to stop.

Model	Handles well	Termination
Flow matching	continuous particle-cloud kinematics	multiplicity supplied externally
Nested-GPT	ordered shower histories	stop/continue learned after each emission

Nested-GPT Architecture

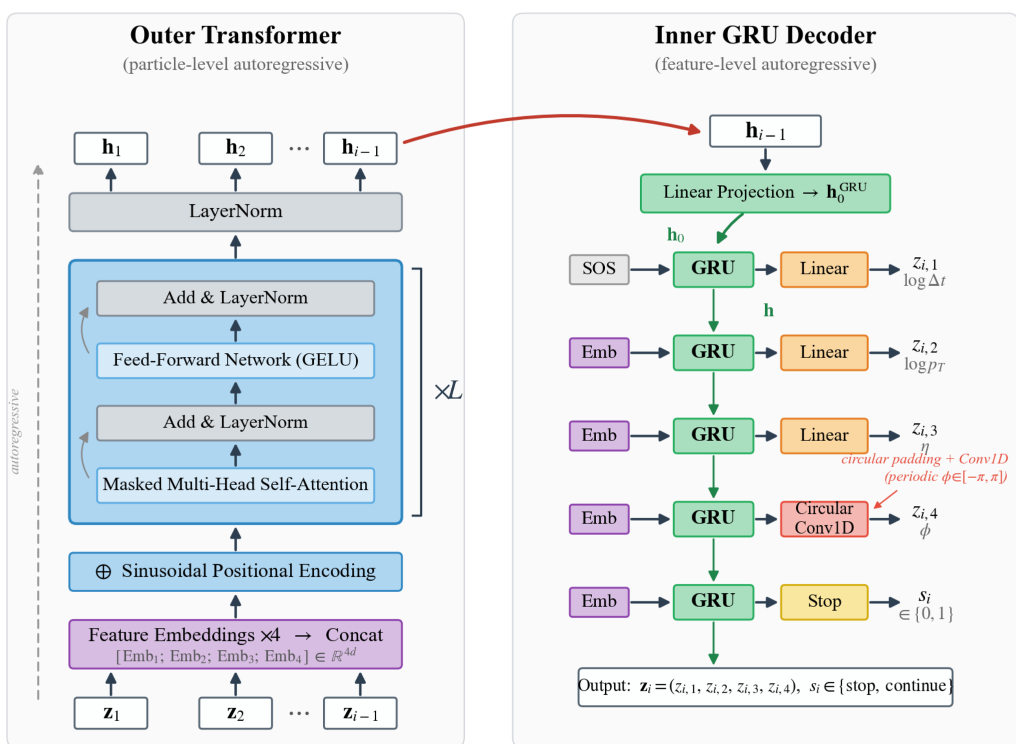


Figure. The outer causal Transformer stores particle-to-particle shower history. The inner GRU decoder generates the next emission feature by feature, then evaluates stop/continue.

- **Outer level:** 6 causal Transformer layers with 8 attention heads encode previous emissions.
- **Inner level:** a 2-layer GRU generates $\log\Delta t, \log p_T, \eta, \phi$ in sequence.
- **Discrete representation:** 256 quantile bins per feature, with ordinal smoothing during training.
- **First emission:** a lightweight leading-particle GRU supplies the learned initial prior.

Nested decoding mirrors the shower hierarchy: event history outside, particle features inside.

Training and Sampling

Step	Sampling procedure
1	sample the first emission from the leading-particle prior
2	autoregressively generate features for the next emission
3	apply the stop/continue head after each generated particle
4	discard events that reach $N_{\max}$ without a stop token

- Teacher forcing predicts the next emission from the ground-truth prefix during training.
- The stop head is trained together with feature logits, so multiplicity is part of the learned distribution.
- Continuous kinematics are recovered by sampling uniformly inside the generated quantile-bin intervals.
- Generation proceeds emission by emission until the stop token is produced.

Training sample	What it tests
Fixed gap, $\lvert y\rvert < 0.8$	Direct learning of veto-triggered termination.
Inclusive, $\lvert y\rvert < 0.01$	Offline veto tests whether shower-time order was learned.

- Observables: gap fraction $G(t)$ and normalized rapidity density $P(y)$.

Direct Fixed-Gap Test

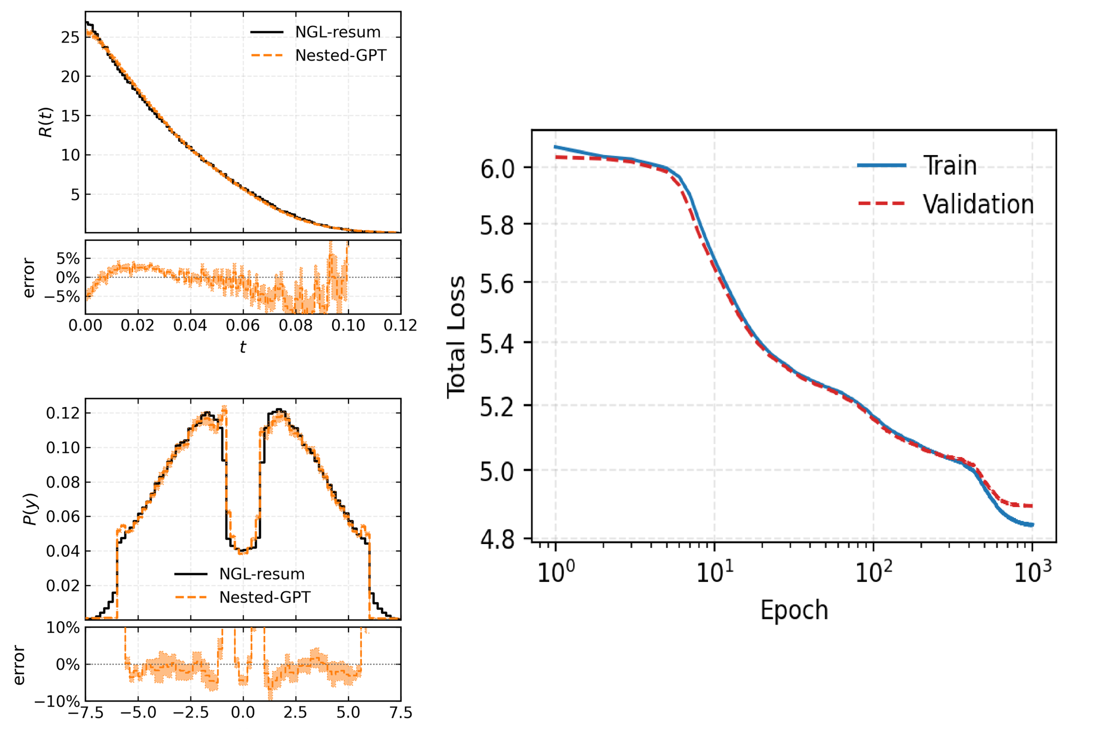


Figure. *Left:* Fixed $\lvert y\rvert < 0.8$ veto; Nested-GPT reproduces $G(t)$ and $P(y)$ against NGL-resum. *Right:* training history; validation loss separates from training loss near epoch 1000.

- The central rapidity depletion and the retained veto-triggering emission are both reproduced.
- This directly tests whether Nested-GPT learned the gap boundary and the physical termination logic.
- Large-t fluctuations arise where the surviving sample is sparse; later epochs show mild overfitting.

Nested-GPT internalizes the phase-space boundary and its sequence-termination logic.

Inclusive-to-Veto Generalization

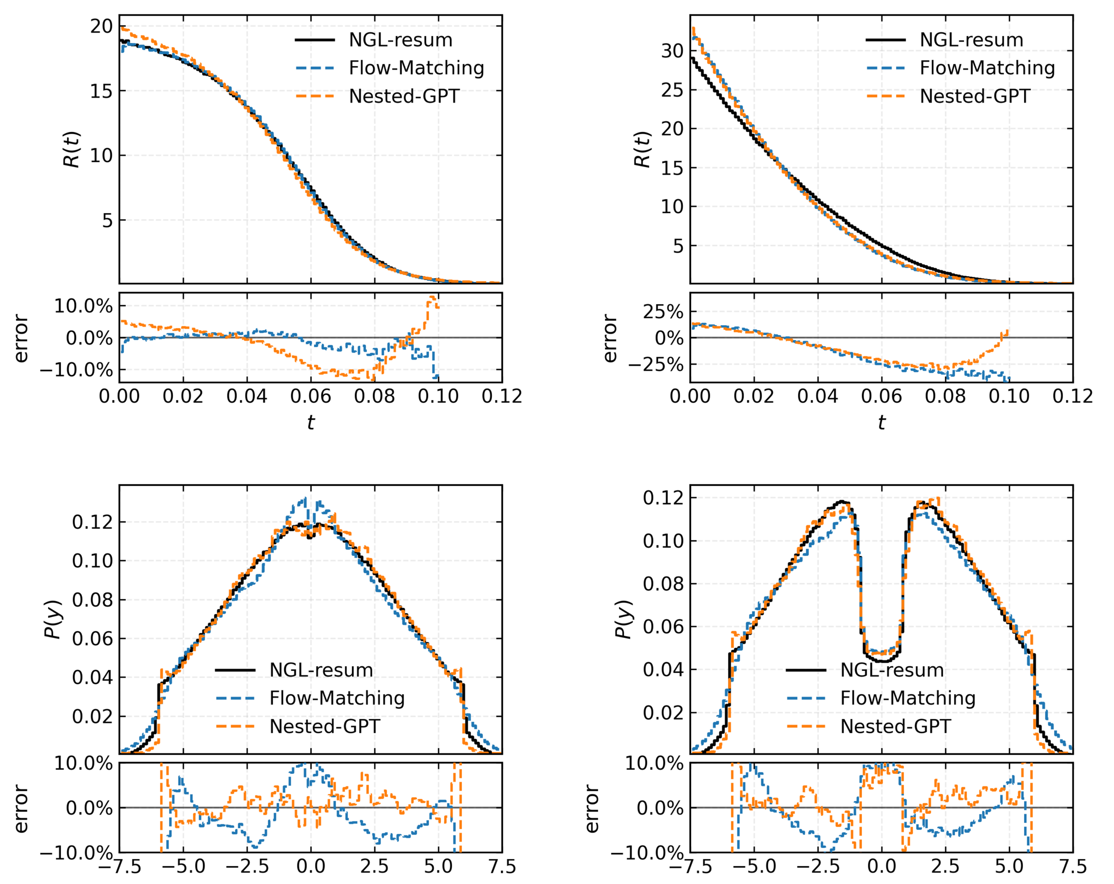


Figure. *Left:* inclusive sample; both flow matching and Nested-GPT match the unrestricted reference shower. *Right:* offline $\lvert y\rvert < 0.8$ truncation; both models recover the vetoed $G(t)$ and $P(y)$.

- The veto is absent during training, so the test probes chronological ordering rather than one-body rapidity fits.
- Both models recover the central depletion and the symmetric two-peak $P(y)$ structure after offline truncation.
- Fidelity shows no significant deterioration relative to direct fixed-gap training for the observables shown.

Main Takeaways

- Parton showers are precision-QCD baselines for Standard Model measurements and new-physics searches.
- NGL resummation is a controlled stress test: nonlinear dipole evolution, ordered branching, and veto-triggered stopping.
- Nested-GPT provides a physically structured surrogate for variable-multiplicity shower histories.
- Flow matching learns continuous kinematics well, but still needs an external multiplicity prior.
- Both architectures reproduce LL large-N_c NGL observables in fixed-gap and offline-veto tests.

Impact	high-statistics surrogate samples for expensive shower evolution
Physics	chronological order, dipole correlations, and veto-triggered termination
Next	finite-N_c color, spin correlations, NLL showers, and realistic collider workflows