UNIVERSITÉ DE GENÈVE

PIPPIN

Generating variable length full events from partons

CHIPP (fast) AI/ML & computing workshop - 19.06.2024

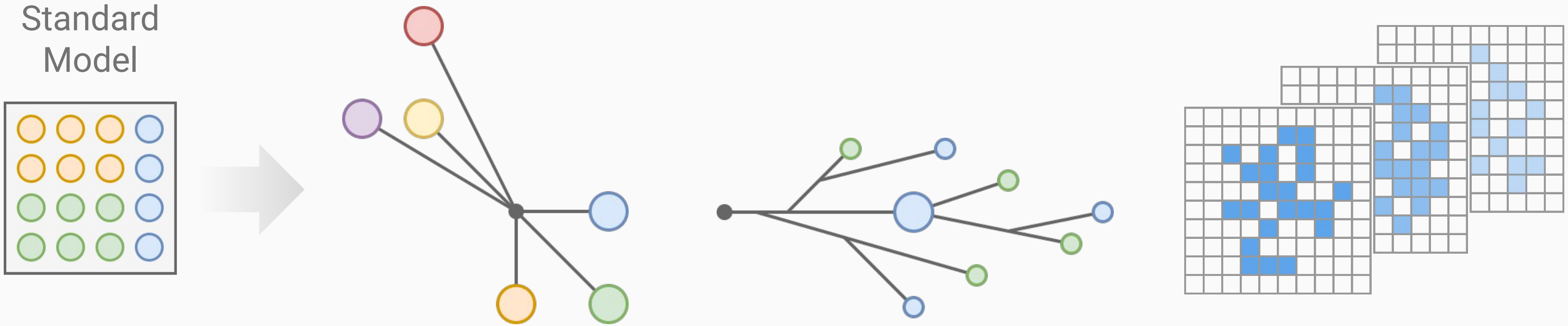
Guillaume Quétant, Johnny Raine, Matthew Leigh, Debajyoti Sengupta, Tobias Golling

Physical process development

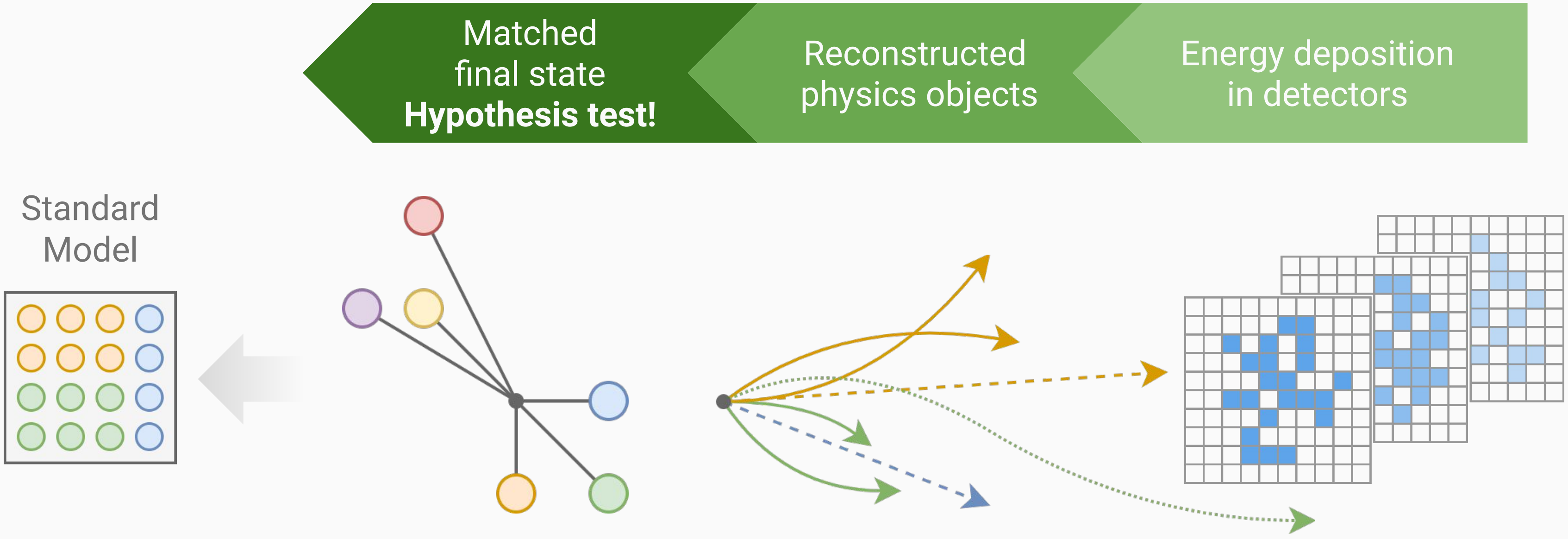
Proton-proton collisions
Physics process!

Particles decay,
Hadronisation,
Showering

Energy deposition
in detectors

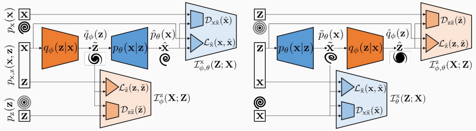


Hypothetical reconstruction

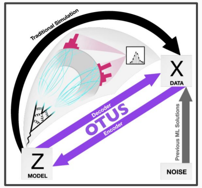


Need simulations

Guillaume Quétant et al.
Turbo-Sim: a generalised generative model with a physical latent space. 2021.
arXiv: 2112.10629



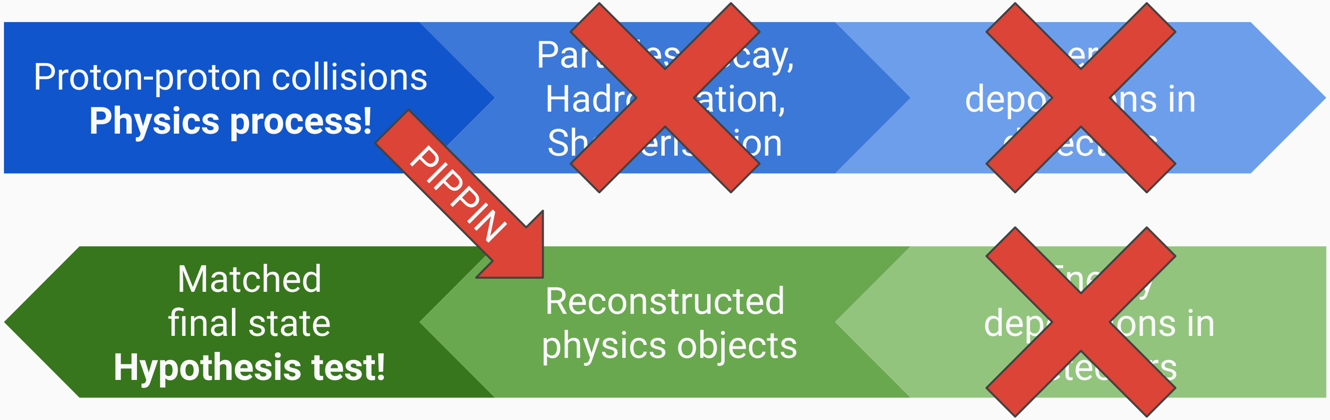
Jessica N. Howard et al.
Learning to simulate high energy particle collisions from unlabeled data. 2021.
Scientific Reports 12, 7567 (2022)



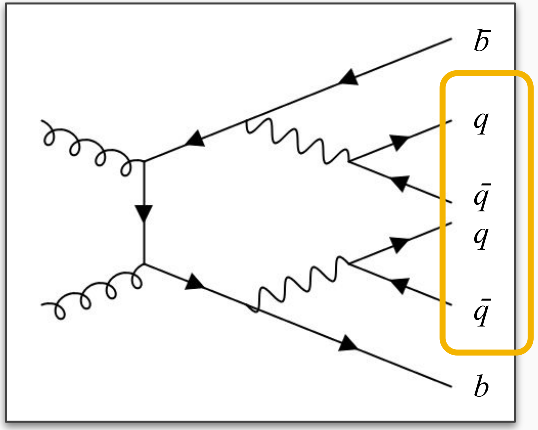
Bypass:

- Decay, showering, hadronisation
- Detector hits and response
- Energy clustering, tracks matching

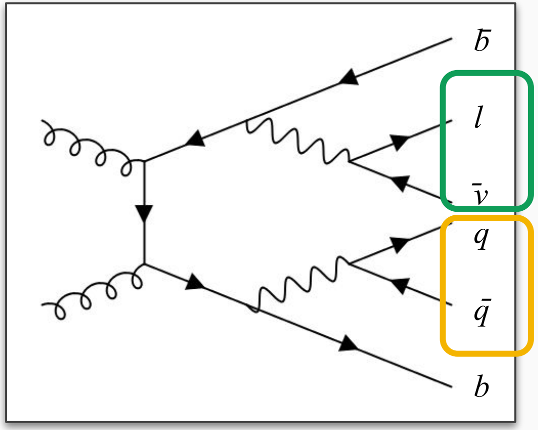
Standard Model



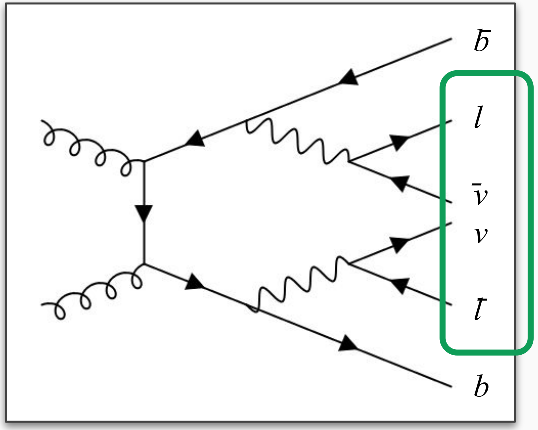
Top quarks study



"all-hadronic"



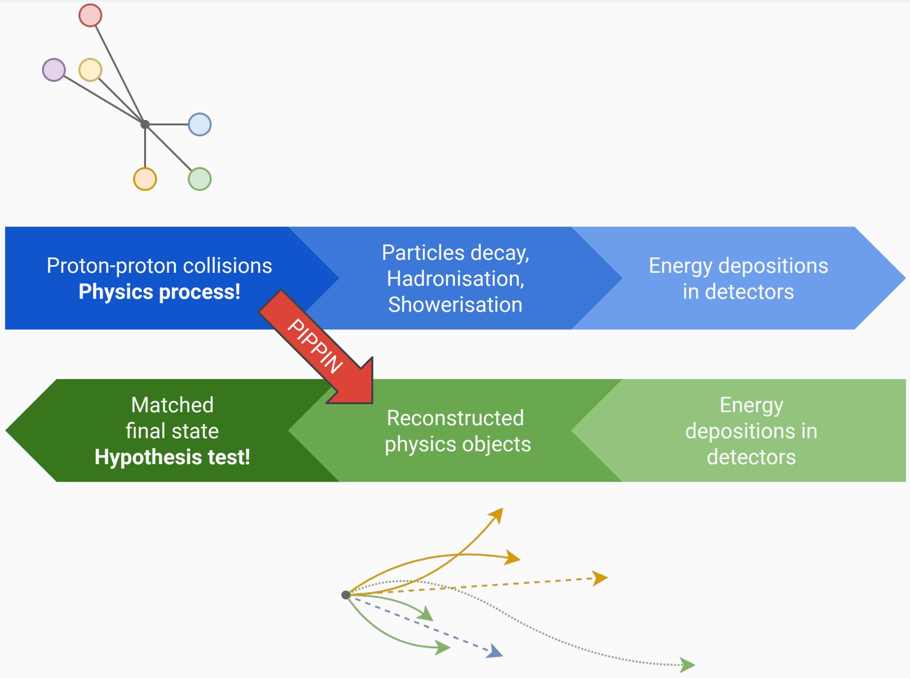
"semi-leptonic"



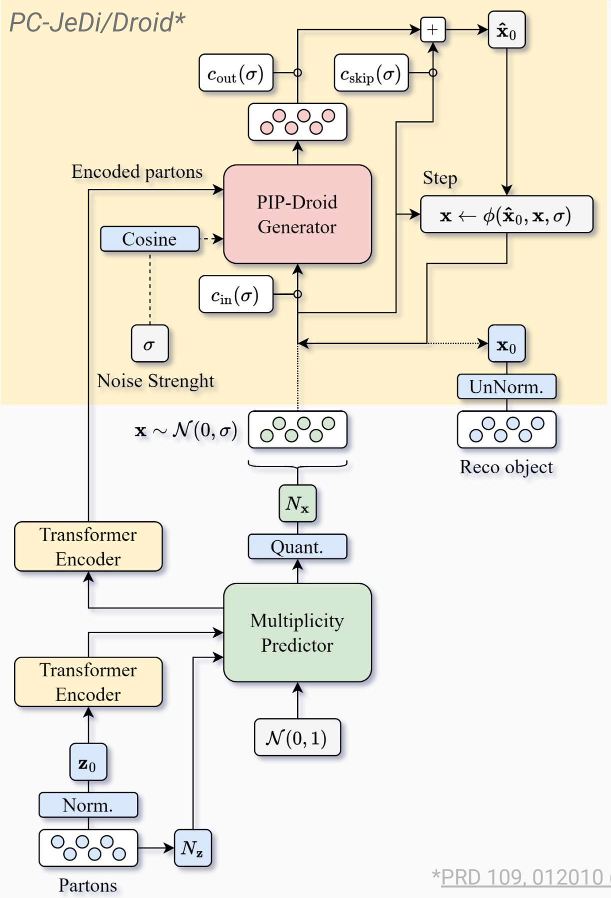
"di-leptonic"

Focus on “partons to reco”

- Inclusive dataset:
  - 18.6M all-hadronic
  - 17.8M semi-leptonic
  - 4.1M di-leptonic
- Input:
  - 6 partons (see previous slide)
- Output:
  - 0-2 leptons
  - MET
  - 2-16 jets
- Features:
  - (p_T, η, φ, E)



Particles Into Particles with Permutation Invariant Network (PIPPIN)

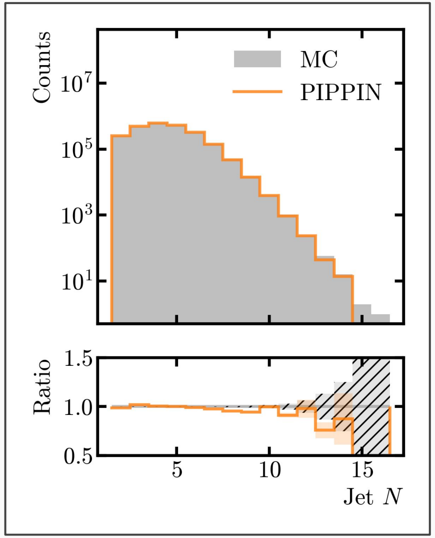


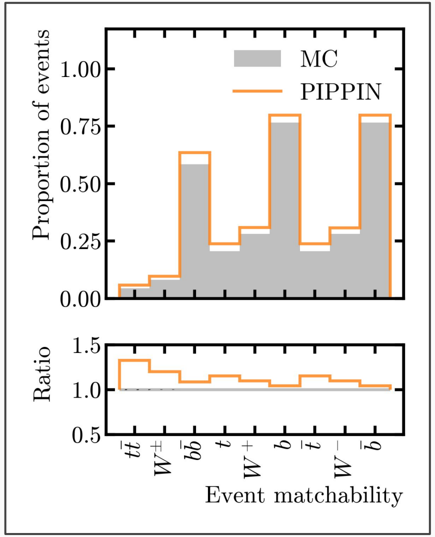
Main characteristics:

- Permutation invariance
  - self/cross-attention
- Conditional networks
- Particle presence
  - auxiliary prediction
- Correlated outputs
  - by attention mechanism
- Stochastic generation
- Three decay channels at once

*[PRD 109, 012010 (2024)]

Multiplicity and presence predictions

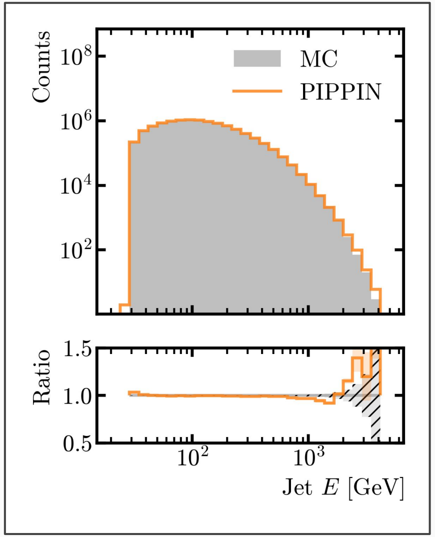


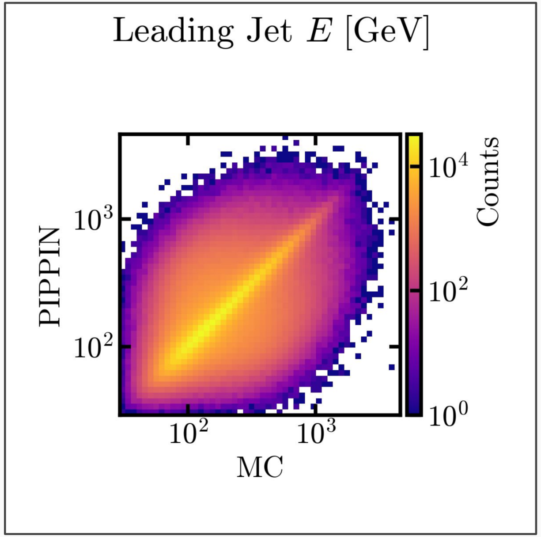


- Very good agreement
- Slight underestimation of number of jets
  - difficulties with low populated tail

- Reasonable prediction
- Slight overestimation in all cases
  - but not a critical quantity to model

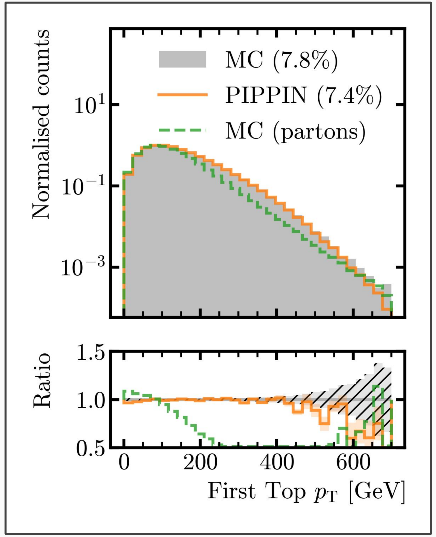
Kinematics properties

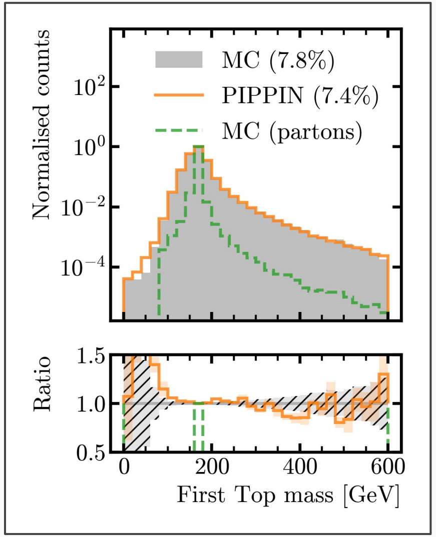




- Very good agreement ← Target of the training
- Struggles a bit with tails and hard cut
- Diagonal well populated
  - with natural intrinsic spread

Underlying intermediate particles



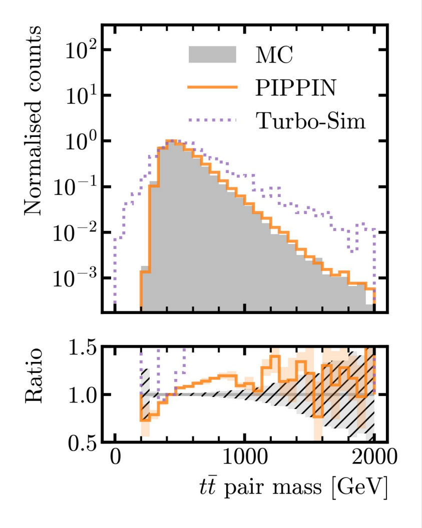


- Similar percentage of fully matched events
  - learnt different topologies w/ same proportion
- Overestimation of low masses
- Underestimation of high masses and momenta

→ Crucial correlations very well handled!

Comparison to similar models

	Reco. objects			Underlying particles			
Model	$p_y^{jet_1}$	$p_z^{jet_1}$	E^{jet_1}	$m^{t\bar{t}}$	m^{W_1}	m^{t_1}	m^{t_2}
OTUS	*3.78*	*2.39*	*5.75*	*15.8*	*11.7*	*14.1*	*24.9*
Turbo-Sim	*2.89*	*10.3*	*4.43*	*2.97*	*7.72*	*5.20*	*8.52*
Turbo-Sim (new)	8.63	12.6	6.32	7.90*	38.8*	38.5*	43.6*
PIPPIN	**0.32**	**0.33**	**0.34**	**4.00**	**3.66**	**3.27**	**2.44**
PIPPIN (inc)	*0.08*	*0.14*	*0.12*	*0.33*	*1.69*	*0.54*	*0.60*



- Comparison on restricted dataset
  - Only semi-leptonic: 1 lepton, MET and 4 jets
- Significantly outperforms other models
- Promising generalisation capabilities

Conclusion

- State-of-the-art partons to variable length full events generative model
  - Correct multiplicities, kinematics and correlations
- Simultaneously learnt three decay channels
  - May extend to more processes!
- Controlled conditional generation

Outlook:

- How well would it handle process variations?
- How to apply the model to the unfolding task?

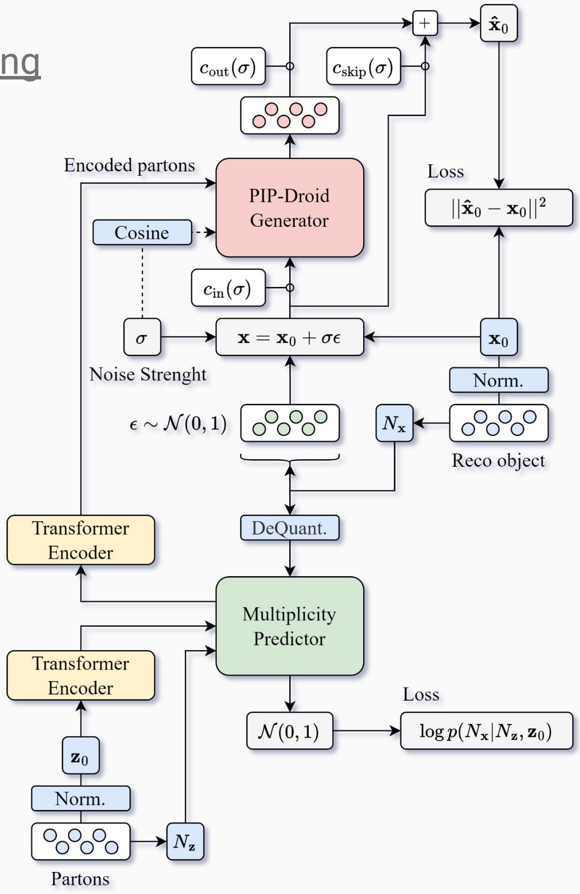
Paper to appear in a couple of days

Thanks!

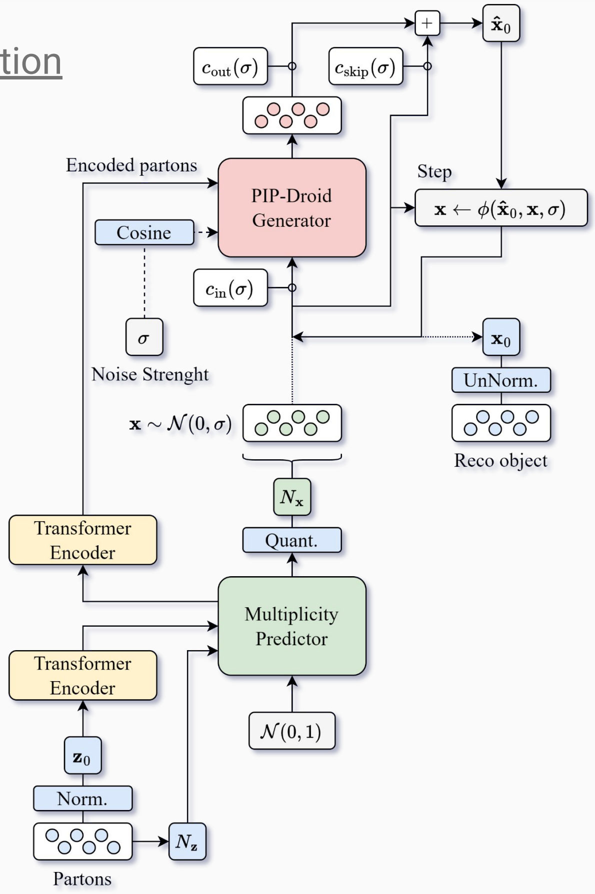
Backup

PIPPIN

Training

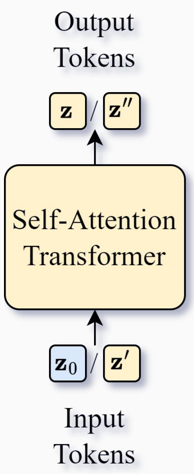


Generation

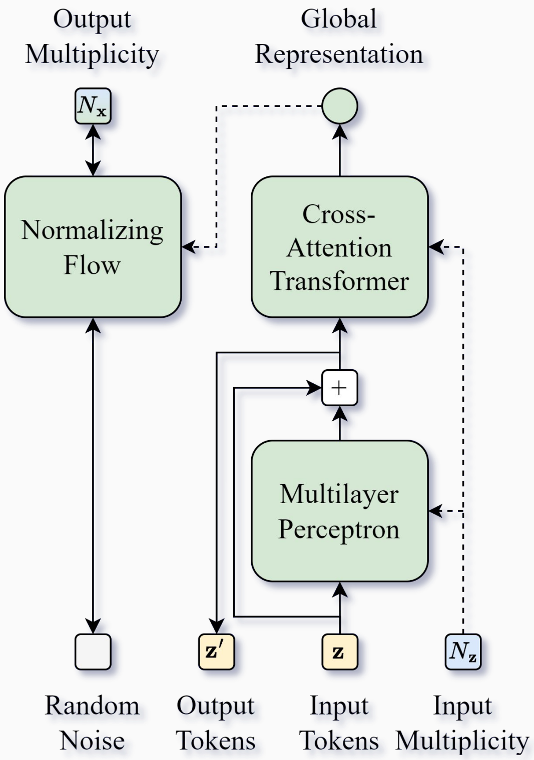


PIPPIN's 3 subparts

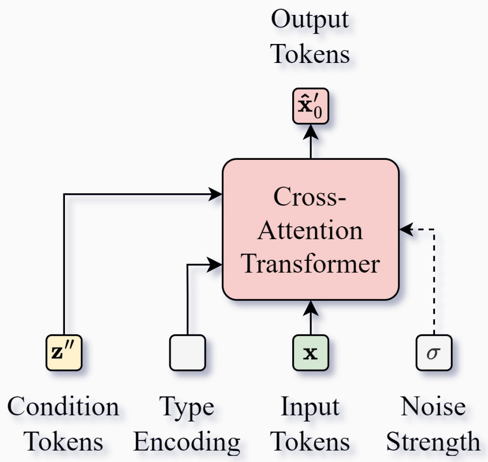
Transformer Encoder



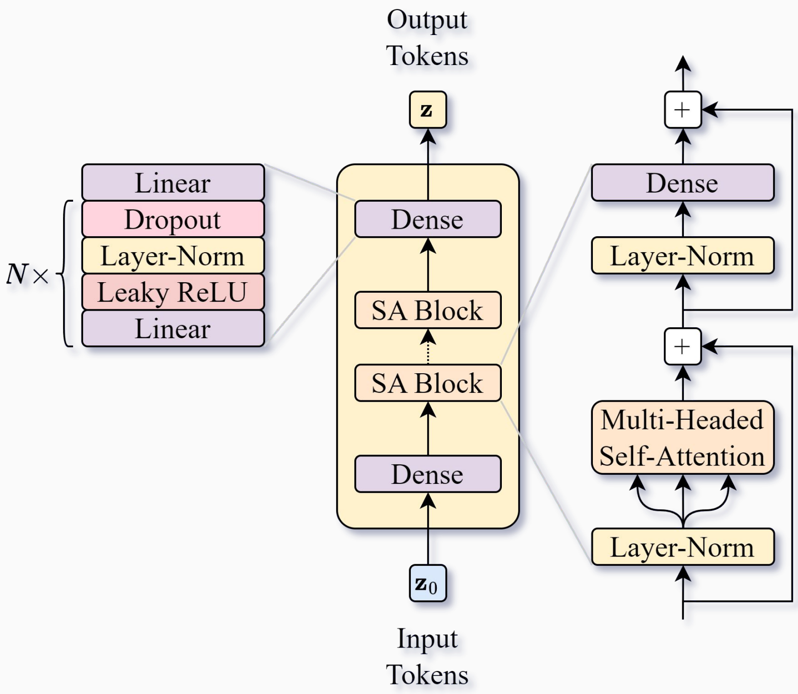
Multiplicity Predictor



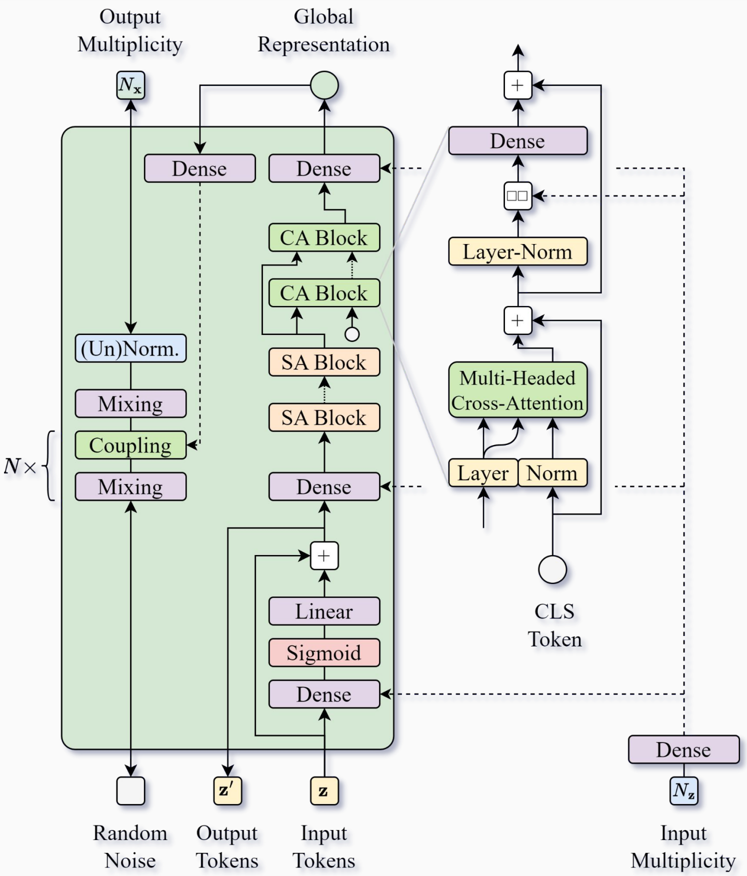
PIP-Droid Generator



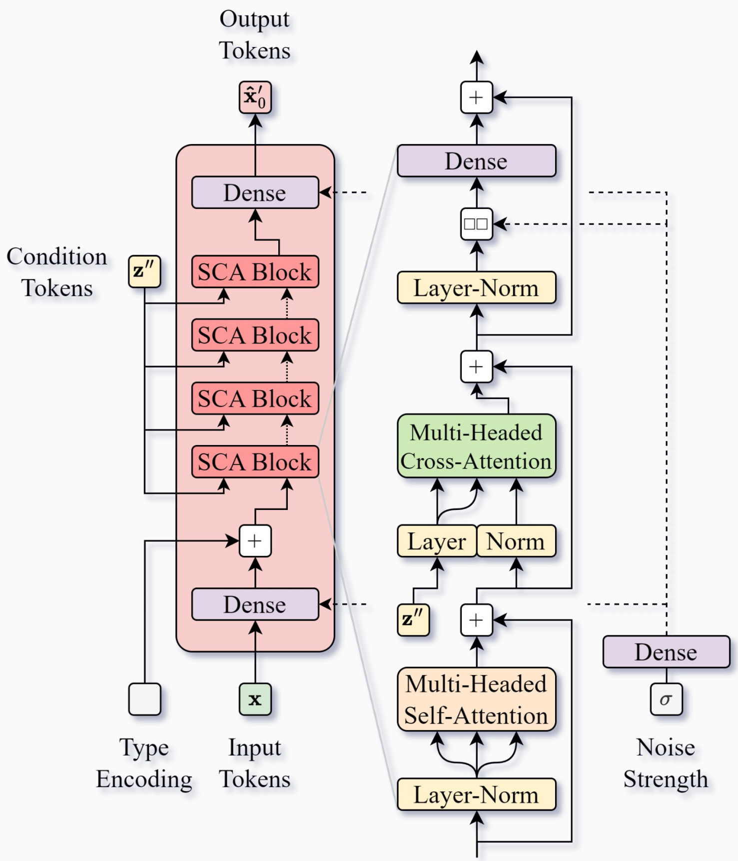
Transformer Encoder



Multiplicity Predictor



PIP-Droid Generator



Comparison of datasets