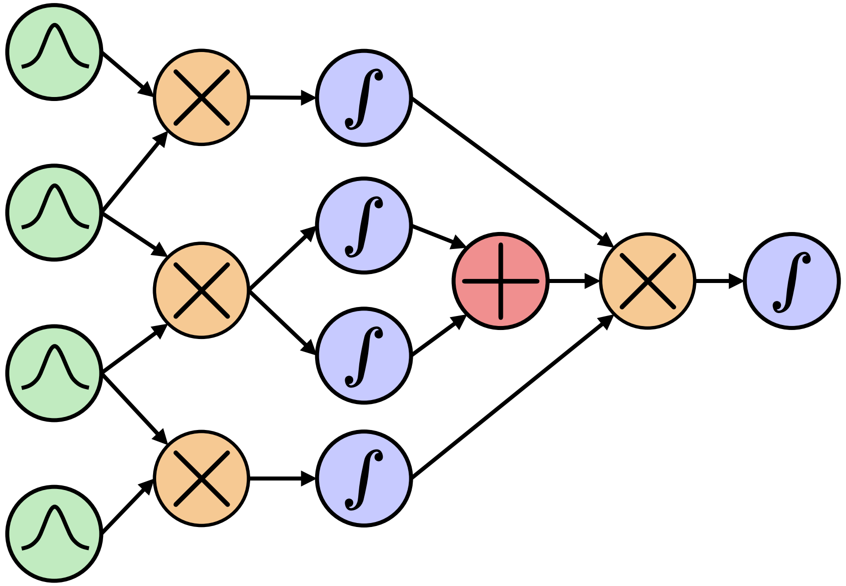
Scaling Continuous Latent Variable Models as Probabilistic Integral Circuits

NEURAL INFORMATION PROCESSING SYSTEMS

TL;DR: We learn continuous hierarchical mixtures as DAG-shaped ***PICs****, and scale them using neural functional sharing techniques.*



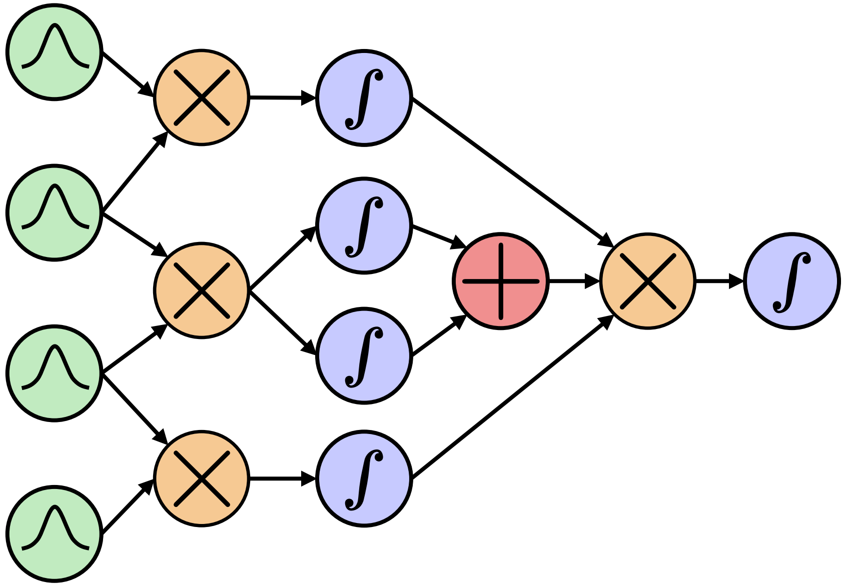
Gennaro Gala[1] Cassio de Campos[1] Antonio Vergari[2] Erik Quaeghebeur[1]

[1] Eindhoven University of Technology, NL

[2] University of Edinburgh, UK

Background – Probabilistic Integral Circuits

- PICs are symbolic computational graphs over possibly non-normalized distributions, and represent hierarchical continuous mixture models using input, product, sum and integral units.
- Non-input units take one or more functions as input and output a single function
- Functions are 'attached' to input and integral units only



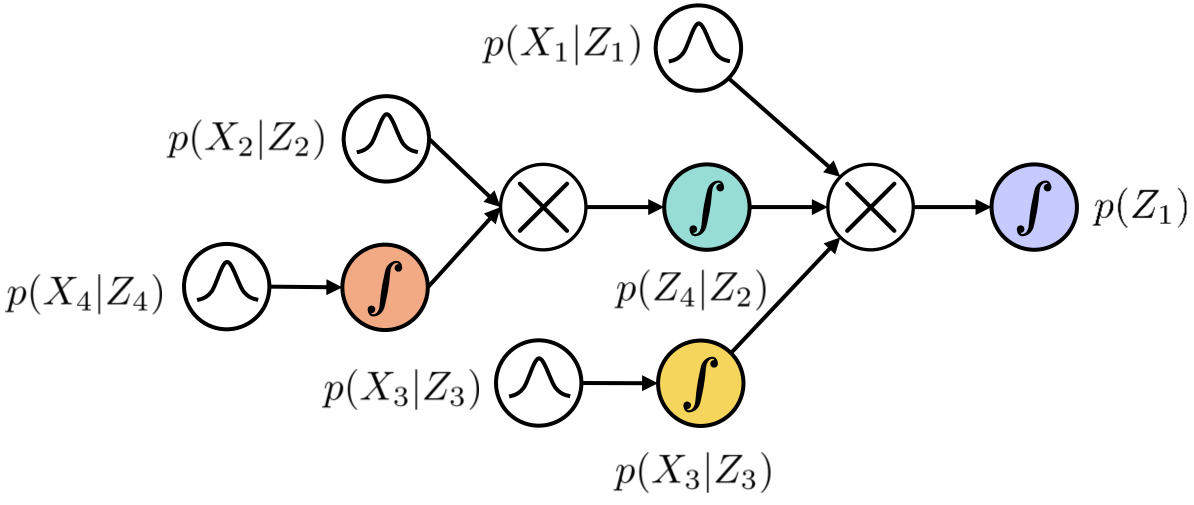
X input variable

Z latent variable

$$f_1(X_1, Z_1) \xrightarrow[f_2(Z_2, Z_1)]{\int} g_2(X_1, Z_2) = \int f_2(Z_2, z_1) f_1(X_1, z_1) \mathrm{d}z_1$$

Previous work & its limitation

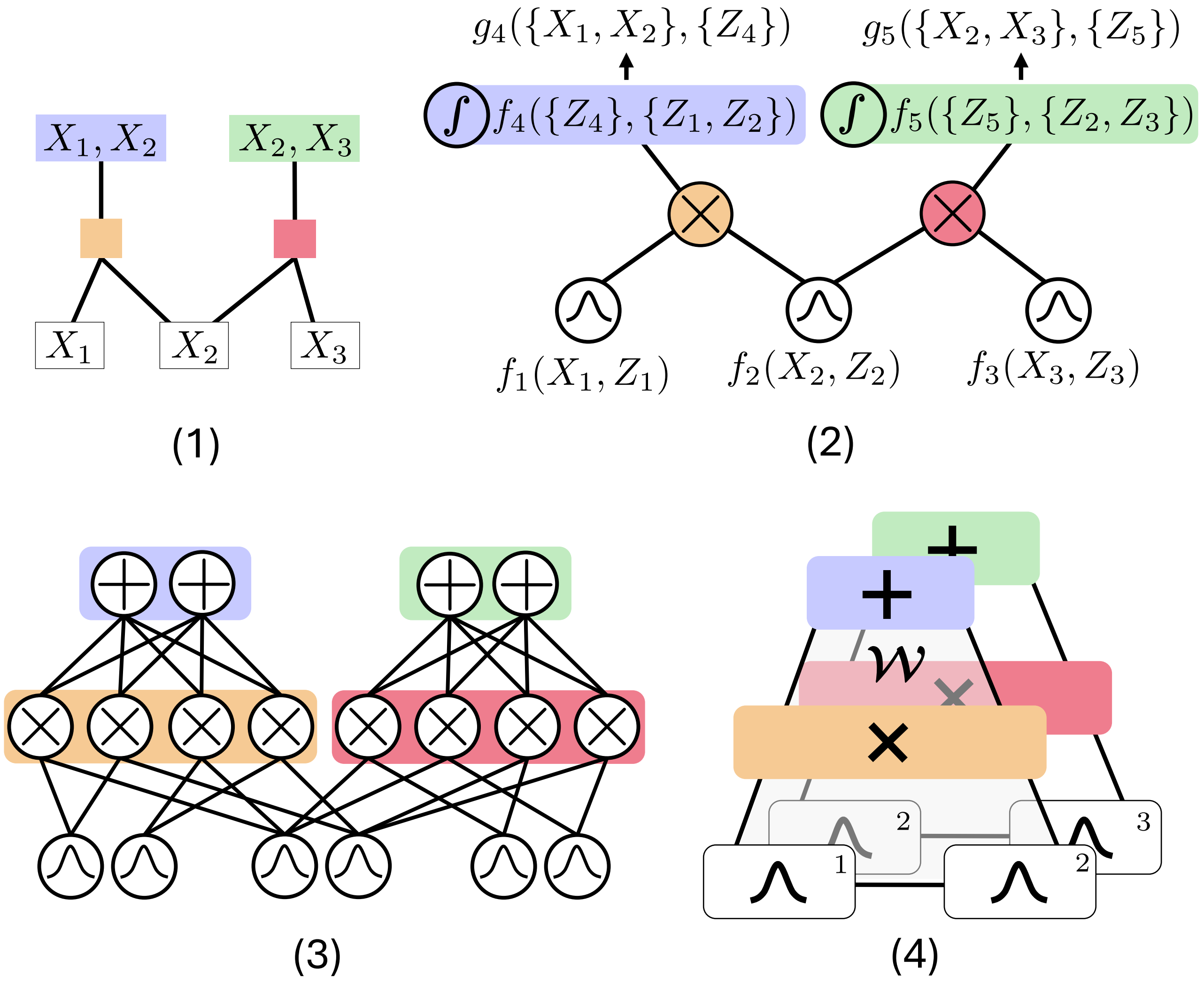
- In previous work [1], PICs where **(i)** limited to tree-shaped structure and **(ii)** only used univariate dependencies between latent variables as to make training feasible
- **RQ**: *How can we build more intricate structures and allow for multivariate latent relationships while providing scalable training?*



$$p(\mathbf{X}) = \int p(z_1) p(X_1|z_1) \int p(z_2|z_1) p(X_2|z_2) \int p(z_4|z_2) p(X_4|z_4) \mathrm{d}z_4 \mathrm{d}z_2 \int p(z_3|z_1) p(X_3|z_3) \mathrm{d}z_3 \mathrm{d}z_1$$

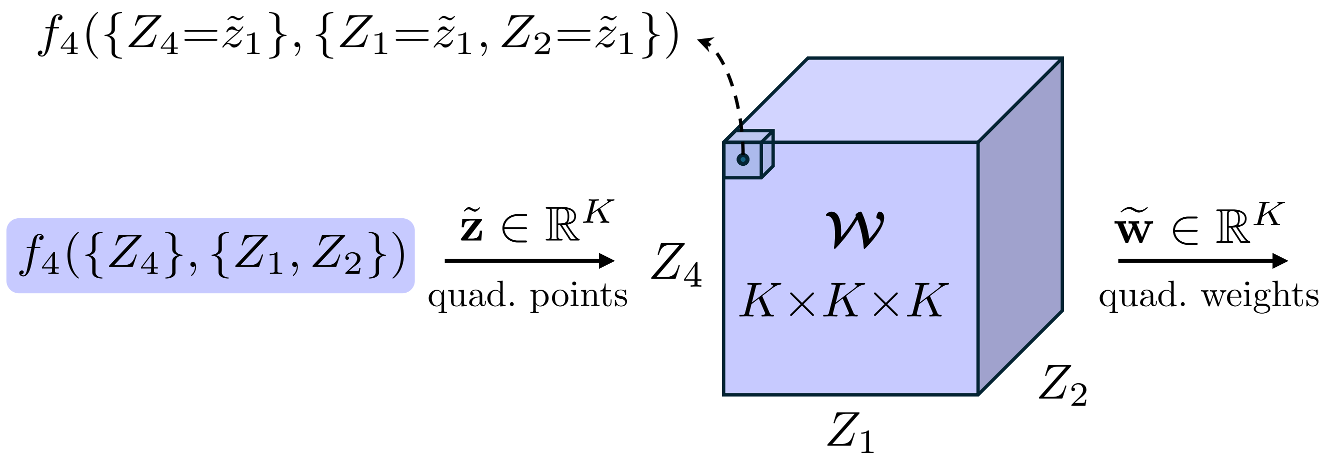
[1] Gala et al. "Probabilistic integral circuits." *AISTATS* 2024.

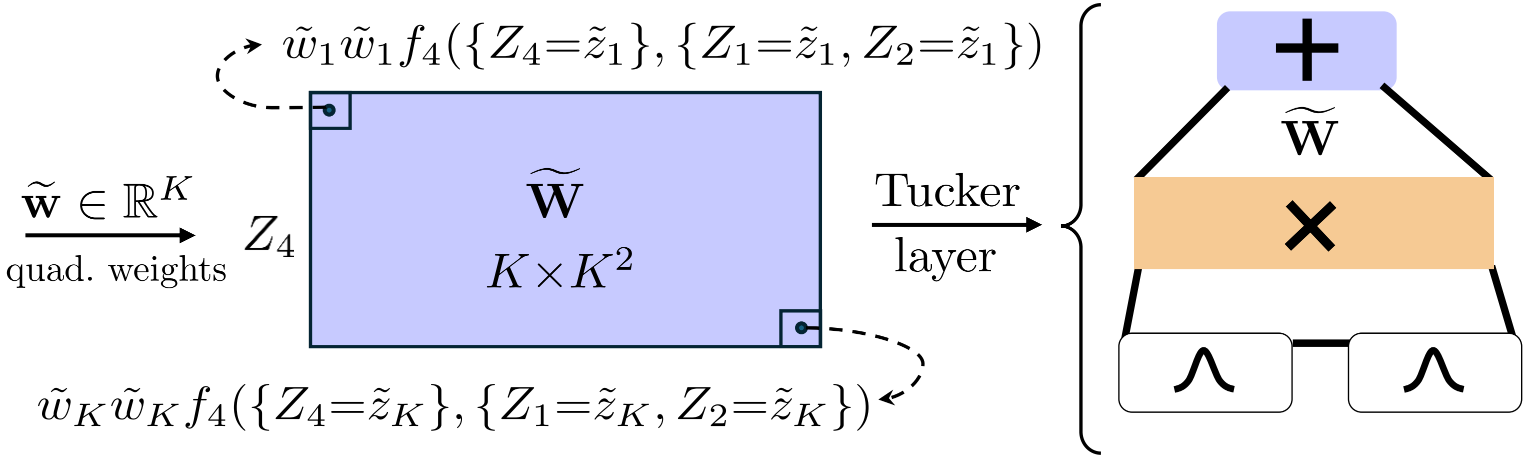
A scalable pipeline to build & learn PICs



We present a pipeline that from arbitrary variable decompositions (1) builds DAG-shaped PICs (2), that we train by materializing them as tensorized circuits (aka *tensor networks*) called Quadrature-PCs (QPCs) (3), which we also fold to allow fast inference (4).

PIC2QPC: The Tucker layer case

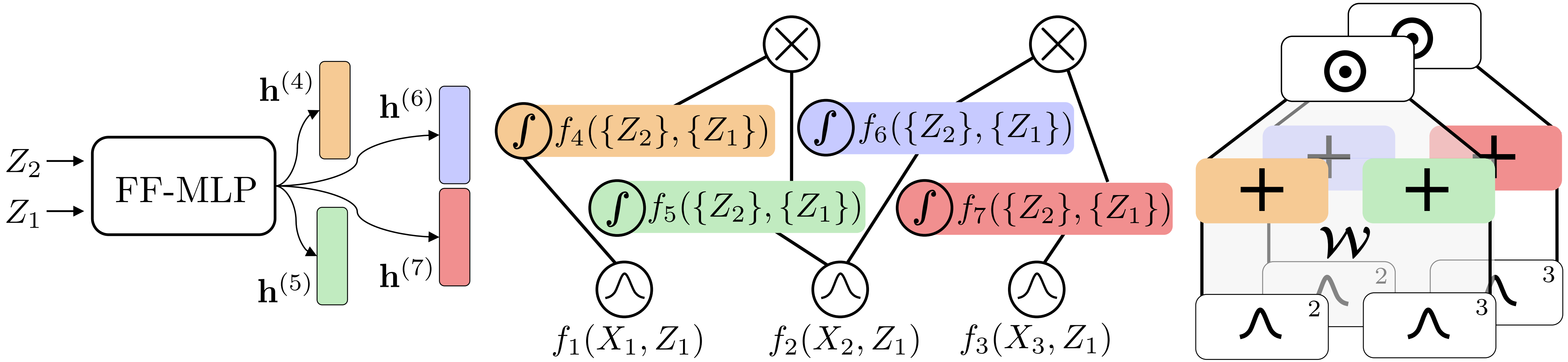




- Zooming-in the QPC materialization, we show how the function f_4 can be discretized via numerical quadrature and used to parameterize a Tucker layer.
- The two gaussian blocks are just vectors of size K, which get multiplied via an outer product that is then matrix-multiplied by $\widetilde{\mathbf{W}}$

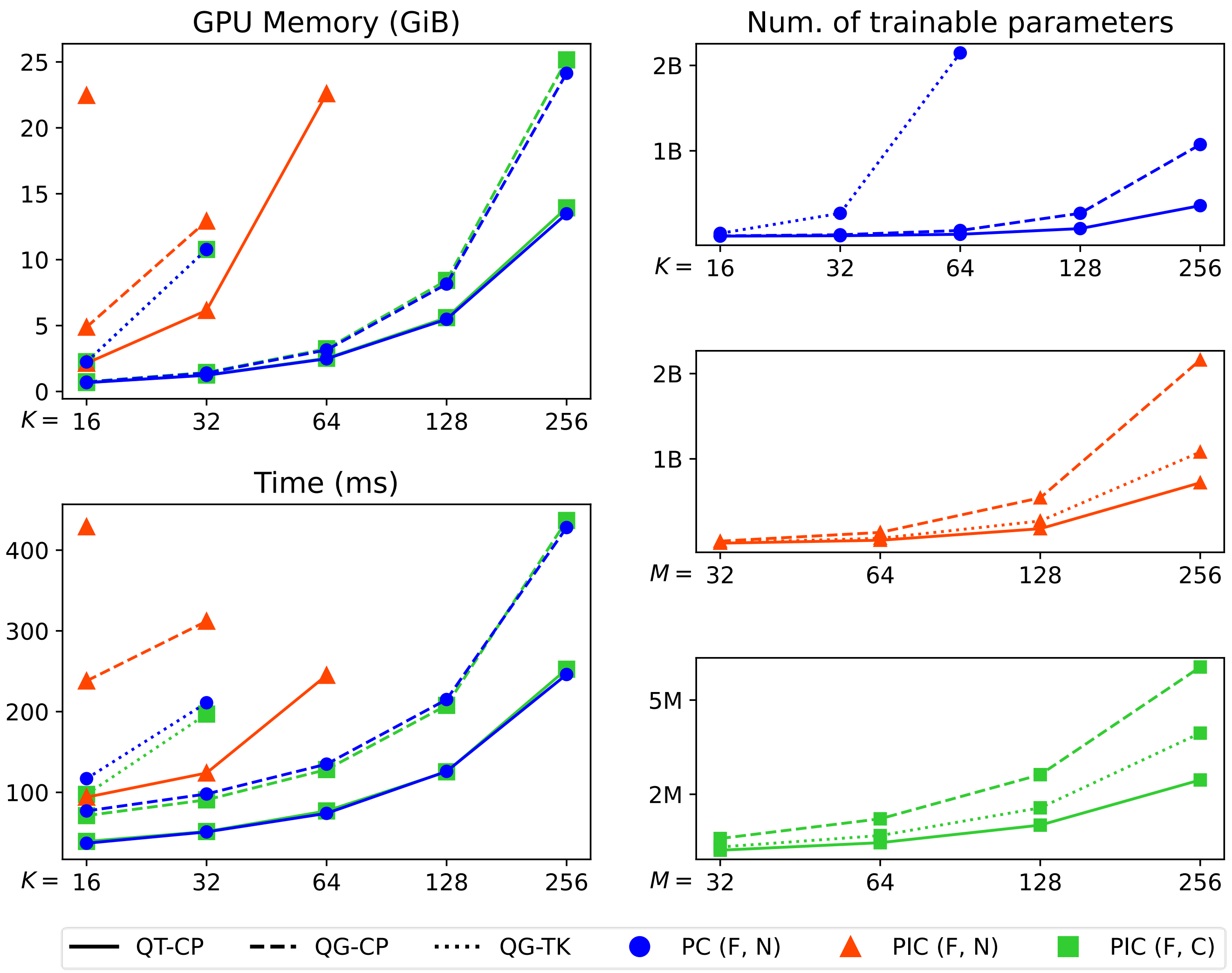
Neural functional sharing for faster & cheaper QPC materialization

- Materializing QPCs is expensive when function evaluation is costly, so we present ***neural functional sharing***: We parameterize all integral units with the same functional form and at the same depth using a multi-headed MLP.



(1) multi-headed MLP (2) DAG-shaped PIC (3) folded CP layer

Neural functional sharing makes PICs scale



PICs with functional sharing (■) - unlike those w/o (▲) - need same resources as PCs (●), and use up to 99% less params!

- QT-CP, QG-CP, QG-TK are tensorized circuit architectures, and K is the width of their layers.
- M is the size of the PIC MLPs.

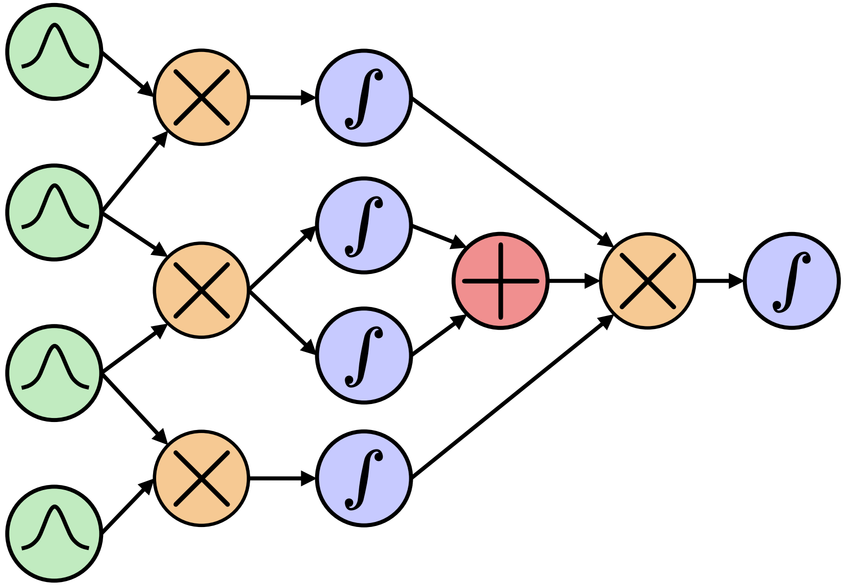
QPCs are performant tractable probabilistic models

	QPC	PC	Sp-PC	HCLT	RAT	IDF	BitS	BBans	McB
MNIST	**1.11**	1.17	1.14	1.21	1.67	1.90	1.27	1.39	1.98
F-MNIST	**3.16**	3.32	3.27	3.34	4.29	3.47	3.28	3.66	3.72
EMN-MN	1.55	1.64	**1.52**	1.70	2.56	2.07	1.88	2.04	2.19
EMN-LE	**1.54**	1.62	1.58	1.75	2.73	1.95	1.84	2.26	3.12
EMN-BA	**1.59**	1.66	1.60	1.78	2.78	2.15	1.96	2.23	2.88
EMN-BY	1.53	**1.47**	1.54	1.73	2.72	1.98	1.87	2.23	3.14
	QPC	PC	HCLT	LVD	LVD-PG				
CIFAR10	**4.48**	4.85	4.61	4.37	3.87				
ImgNet32	**4.46**	4.63	4.82	4.38	4.06				
ImgNet64	**4.42**	4.59	4.67	4.12	3.80				
CelebA	**4.11**	4.16	-	-	-				

QPCs outperform standard PCs on distribution estimation benchmarks

Scaling Continuous Latent Variable Models as Probabilistic Integral Circuits

Gennaro Gala, Cassio de Campos, Antonio Vergari, Erik Quaeghebeur



[1] Gala, Gennaro, et al. "Probabilistic integral circuits." *International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.
[2] Loconte, Lorenzo, et al. "What is the Relationship between Tensor Factorizations and Circuits (and How Can We Exploit it)?." *arXiv preprint arXiv:2409.07953* (2024).
[3] Correia, Alvaro HC, et al. "Continuous mixtures of tractable probabilistic models." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. No. 6. 2023.