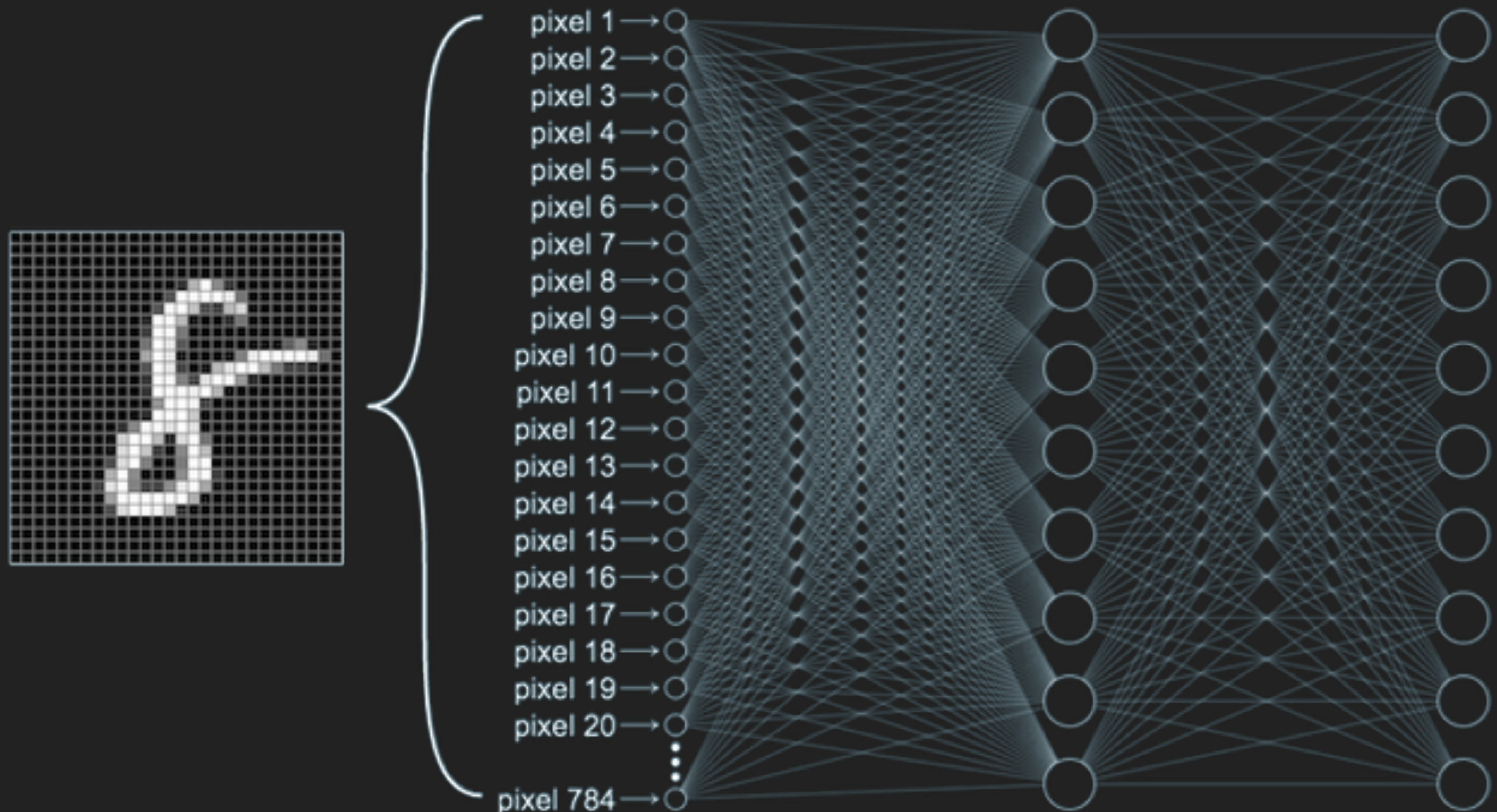
CLARIANT

LORENTZ COVARIANT NEURAL NETWORK

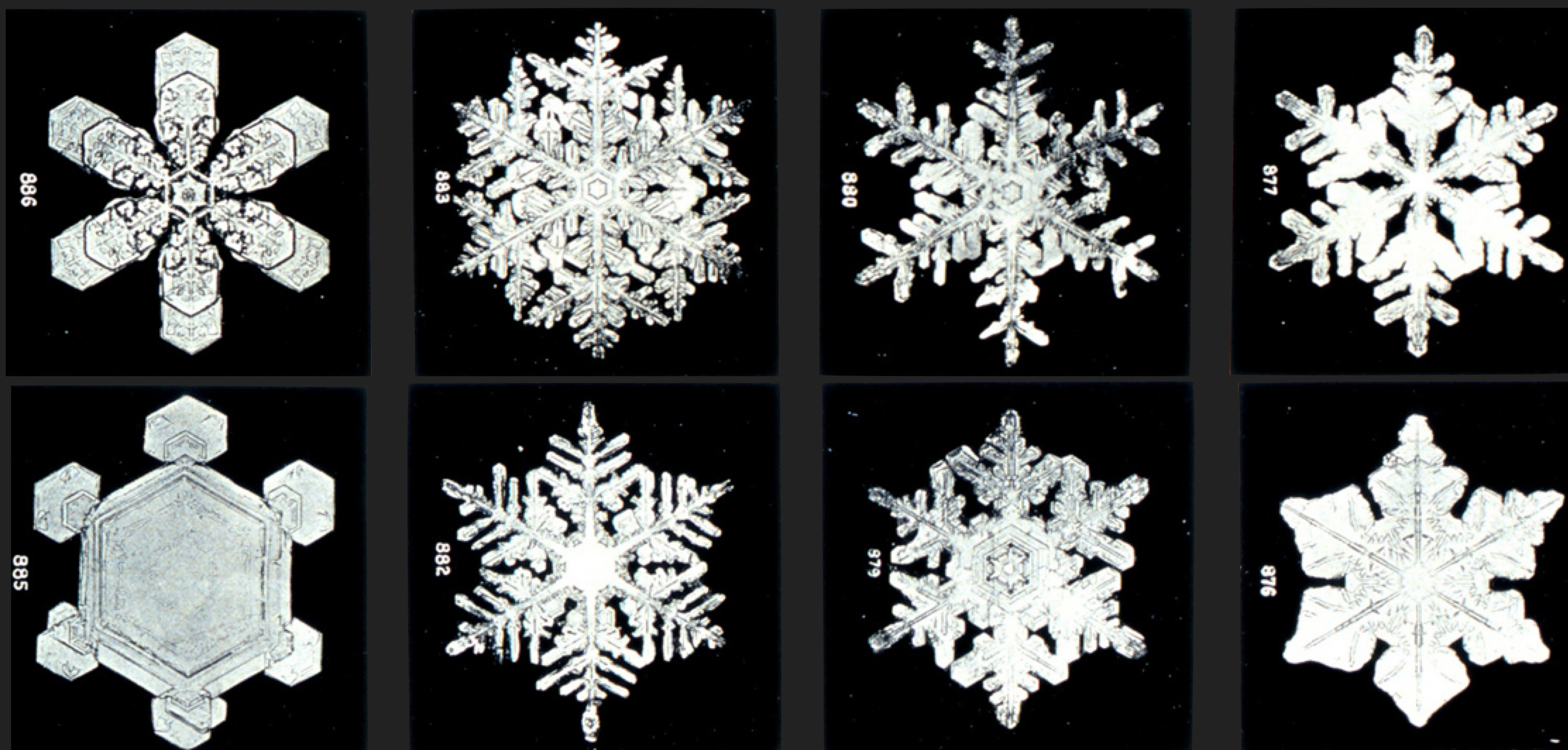
Alexander Bogatskiy, Brandon Anderson, Risi Kondor, David Miller, Jan Offermann, Marwah Roussi

THE UNIVERSITY OF CHICAGO

FLATIRON INSTITUTE



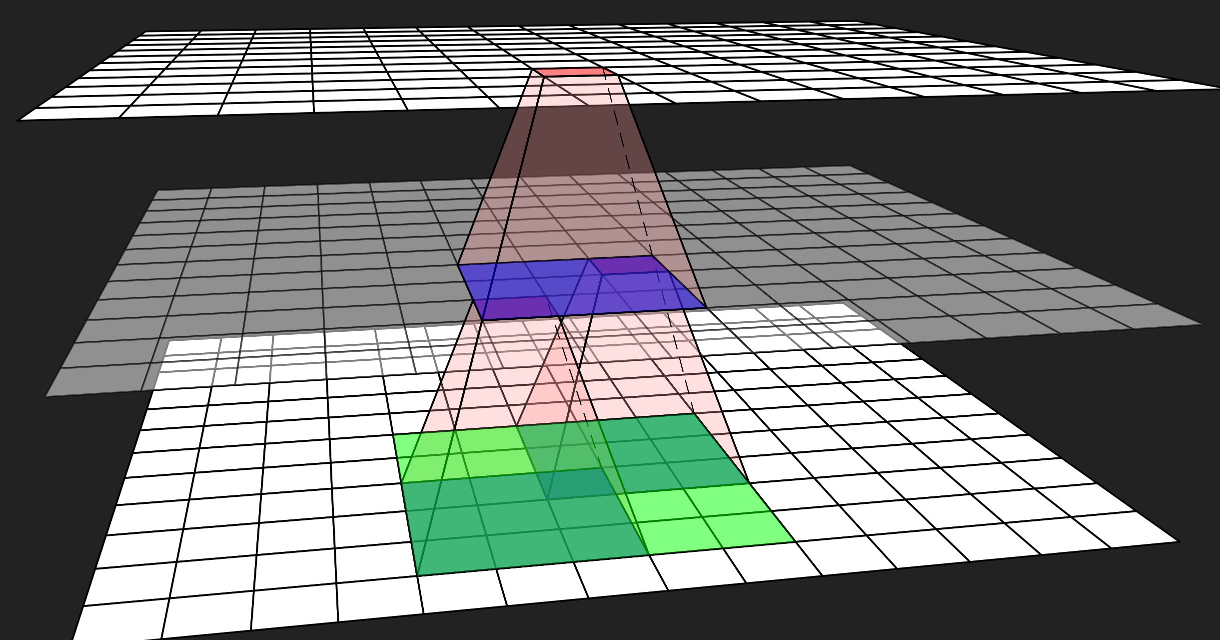
[ml4a.github.io]



[NOAA]

- Solution: CNN’s

- Solution: CNN's



[Cohen & Welling: Group equivariant CNNs (ICML 2016)]

- Solution: randomly rotated training samples?

  preprocessing?

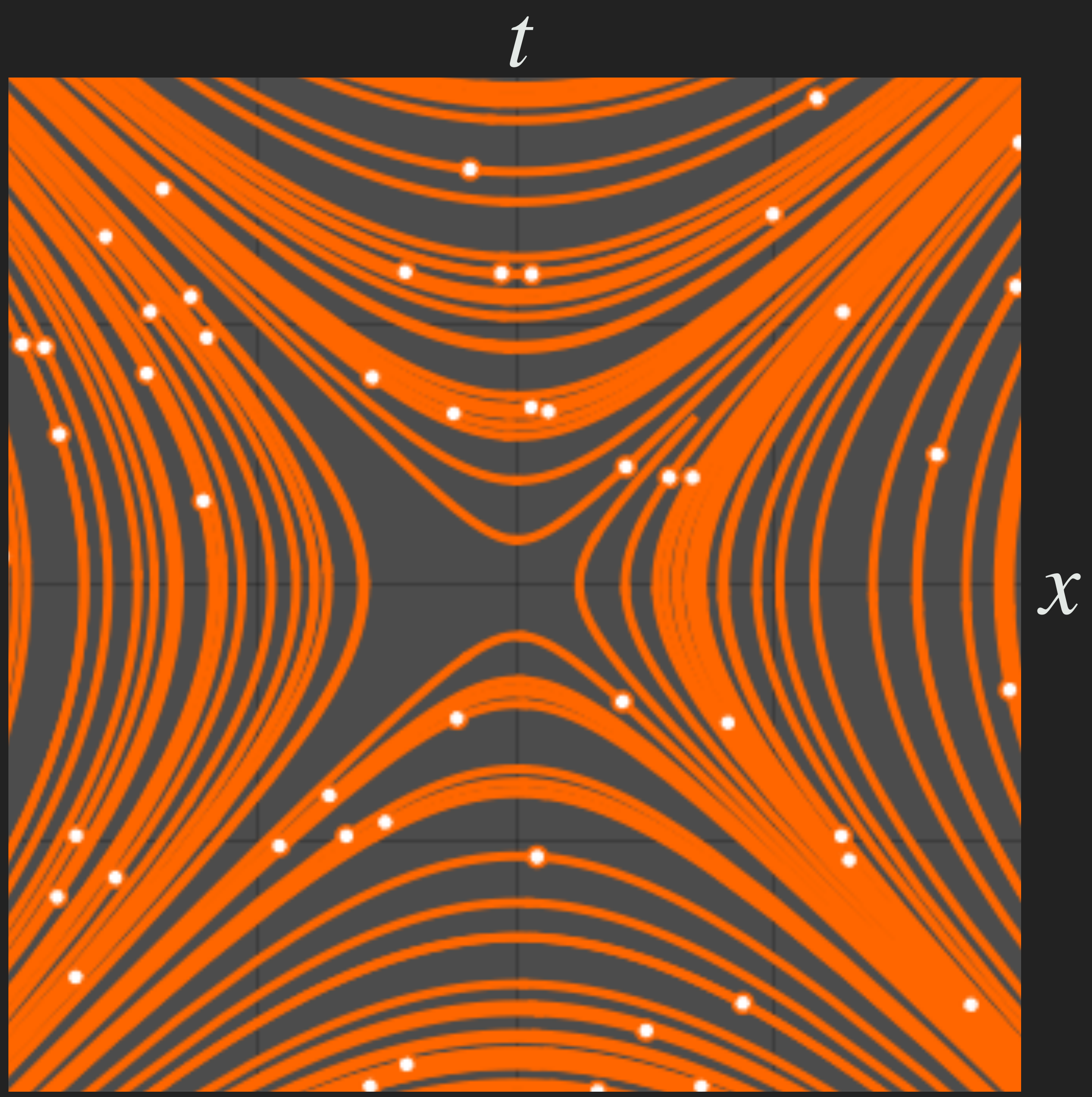
- Solution: ~~randomly rotated training samples?~~

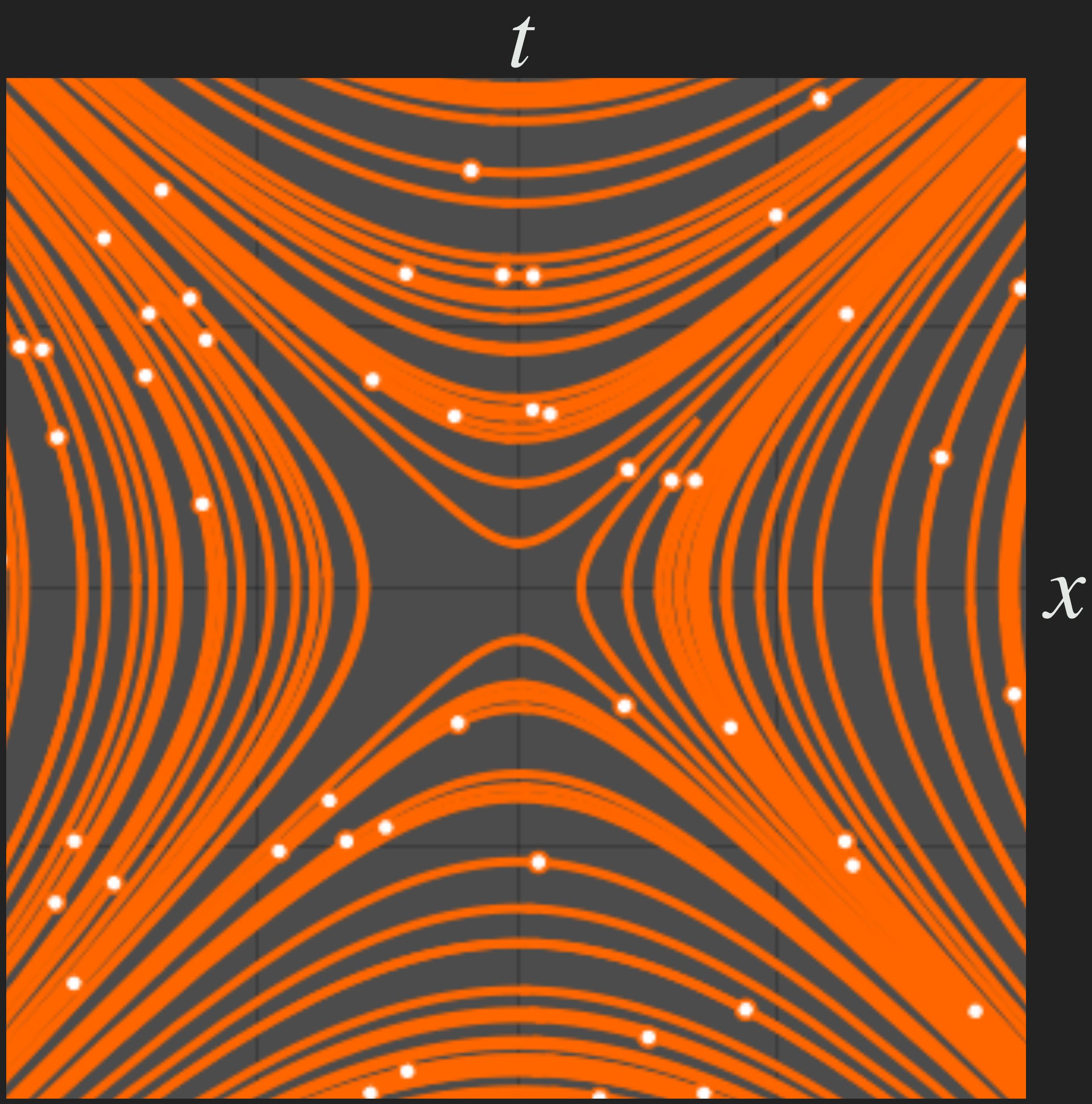
  ~~preprocessing?~~

- Solution: convolutions on Lie groups

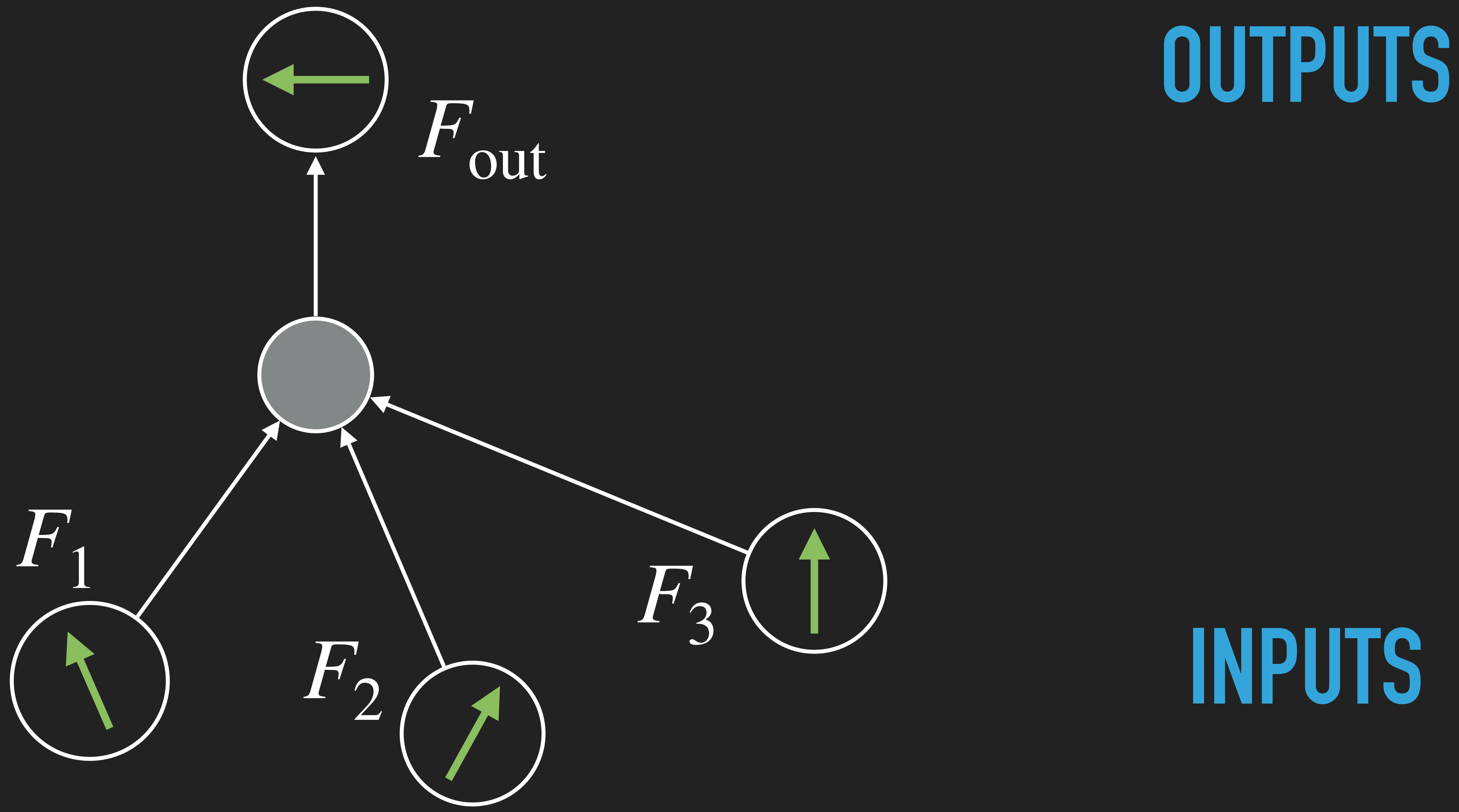
[Cohen & Welling: Group equivariant CNNs (ICML 2016)]
[Cohen & Welling: Steerable CNNs (ICLR 2017)]

- Greatly reduce the number of learnable parameters
- Reduce the number of training samples
- Improve generalization and regression ability
- Provide physically interpretable models
- Bring elegance and mathematical transparency
- Build on physical principles of symmetries and constraints


t
x


t
x

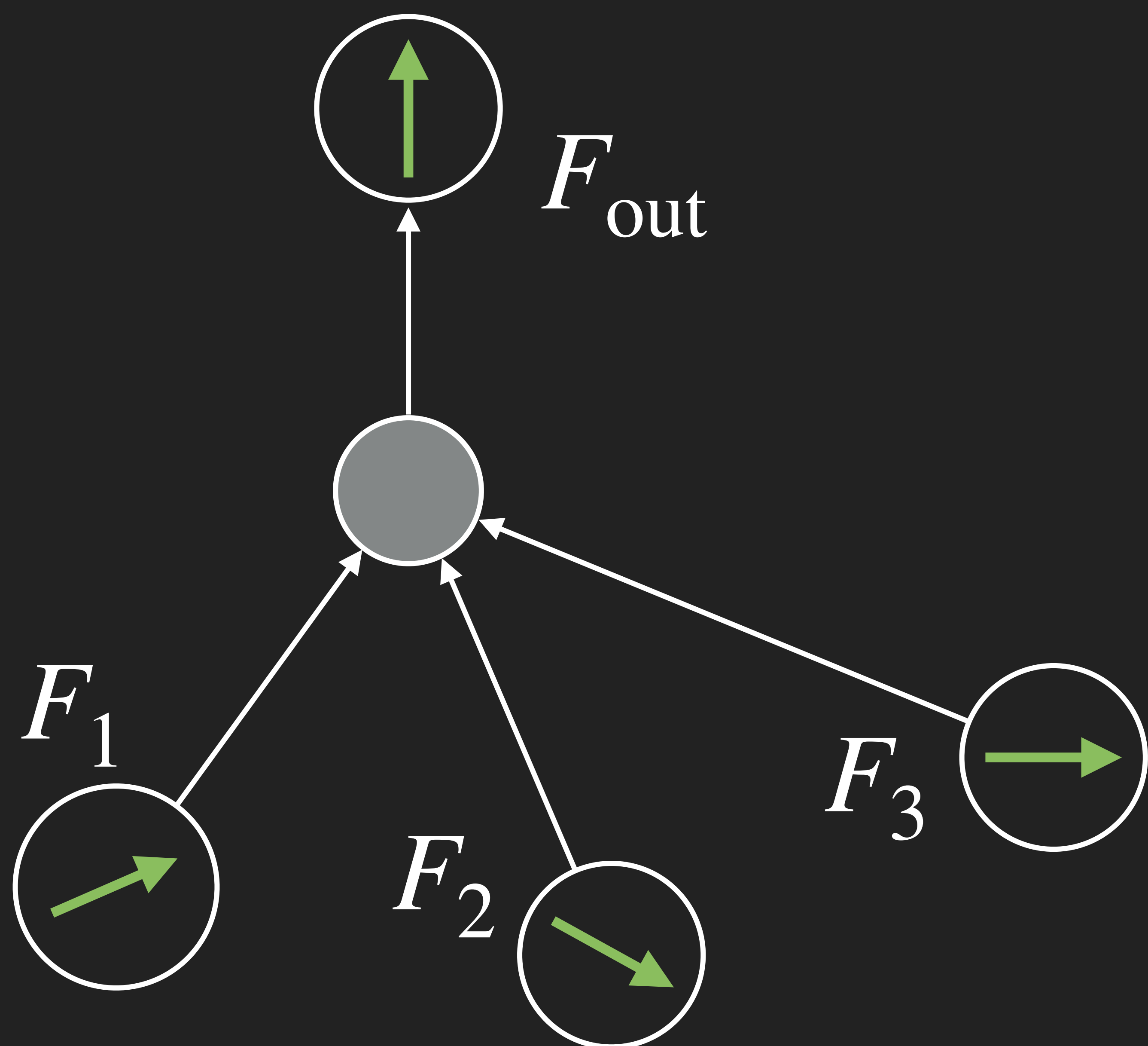
$$F_{\text{out}} \to \rho_{\text{out}}(g) F_{\text{out}}$$



$F_1 \to \rho_1(g) F_1$ $\quad$ $F_2 \to \rho_2(g) F_2$ $\quad$ $F_3 \to \rho_3(g) F_3$

$$F_{\text{out}} \to \rho_{\text{out}}(g)F_{\text{out}}$$

OUTPUTS



INPUTS

$$F_1 \to \rho_1(g)F_1 \qquad F_2 \to \rho_2(g)F_2 \qquad F_3 \to \rho_3(g)F_3$$

Any continuous function can be represented:

$$F(x_1, \ldots, x_n) = \sum_{k=0}^{2n \cdot d} F_k \left(\sum_{i=1}^{n} f_{ki}(x_i) \right)$$

[Kolmogorov-Arnold representation theorem]

Any continuous function can be represented:

$$F(x_1, \ldots, x_n) = \sum_{k=0}^{2n \cdot d} F_k \left(\sum_{i=1}^{n} f_{ki}(x_i) \right)$$

[Kolmogorov-Arnold representation theorem]

Symmetric functions:

$$F(x_1, \ldots, x_n) = \hat{F} \left(\sum_{i} f(x_i) \right)$$

[Zaheer et al. (Deep Sets, 2017)]

Group equivariant neural networks (Cohen & Welling, 2016)

Harmonic networks: deep translation and rotation equivariance (Worrall, Garbin, Turmukhanbetov & Brostow, 2016)

Steerable CNNs (Cohen & Welling, 2017)

On the generalization of convolution and equivariance (Kondor & Trivedi, 2018)

Intertwiners between induced representations (Cohen, Geiger & Weiler, 2018)

3D steerable neural networks (Weiler, Geiger, Wellig, Boomsma & Cohen, 2018)

Gauge equivariant neural networks (Cohen, Weiler, Kicanaoglu, Welling, 2019)

Tensor field networks (Thomas, Smidt, Kearns, Yang, Li Kohlhoff & Riley, 2018)

Relativistic Harmonic Networks (Chase Shimmin @ Boost 2019)

Cormorant: Covariant Molecular Neural Networks (Anderson, Hy & Kondor, 2019)

INGREDIENTS

INGREDIENTS

- Activations valued in a vector space

INGREDIENTS

- Activations valued in a vector space
- learnable linear operation

INGREDIENTS

- Activations valued in a vector space
- learnable linear operation
- composable nonlinear operation

INGREDIENTS

- Activations valued in vector representations of a group
- Equivariant learnable linear operation
- Equivariant composable nonlinear operation

INGREDIENTS

- Activations valued in vector representations of a group
- Equivariant learnable linear operation

  ("commutes" with the action of the group)
- Equivariant composable nonlinear operation

INGREDIENTS

- Activations valued in vector representations of a group
- Equivariant learnable linear operation

  ("commutes" with the action of the group)
- Equivariant composable nonlinear operation

  (maps representations to representations)

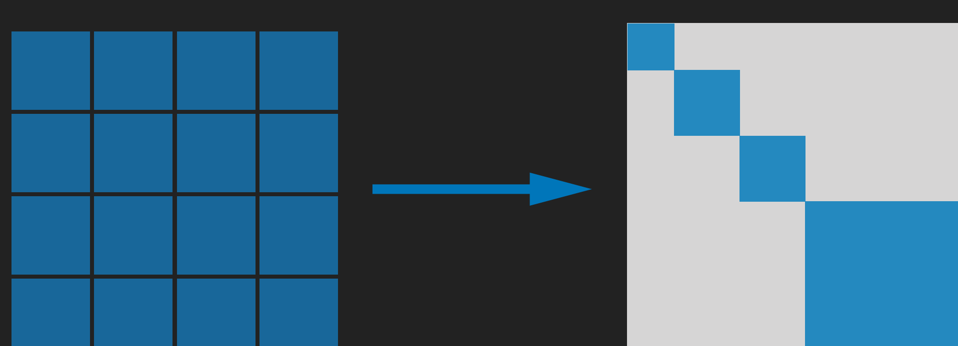
I. MIXING IRREDUCIBLES

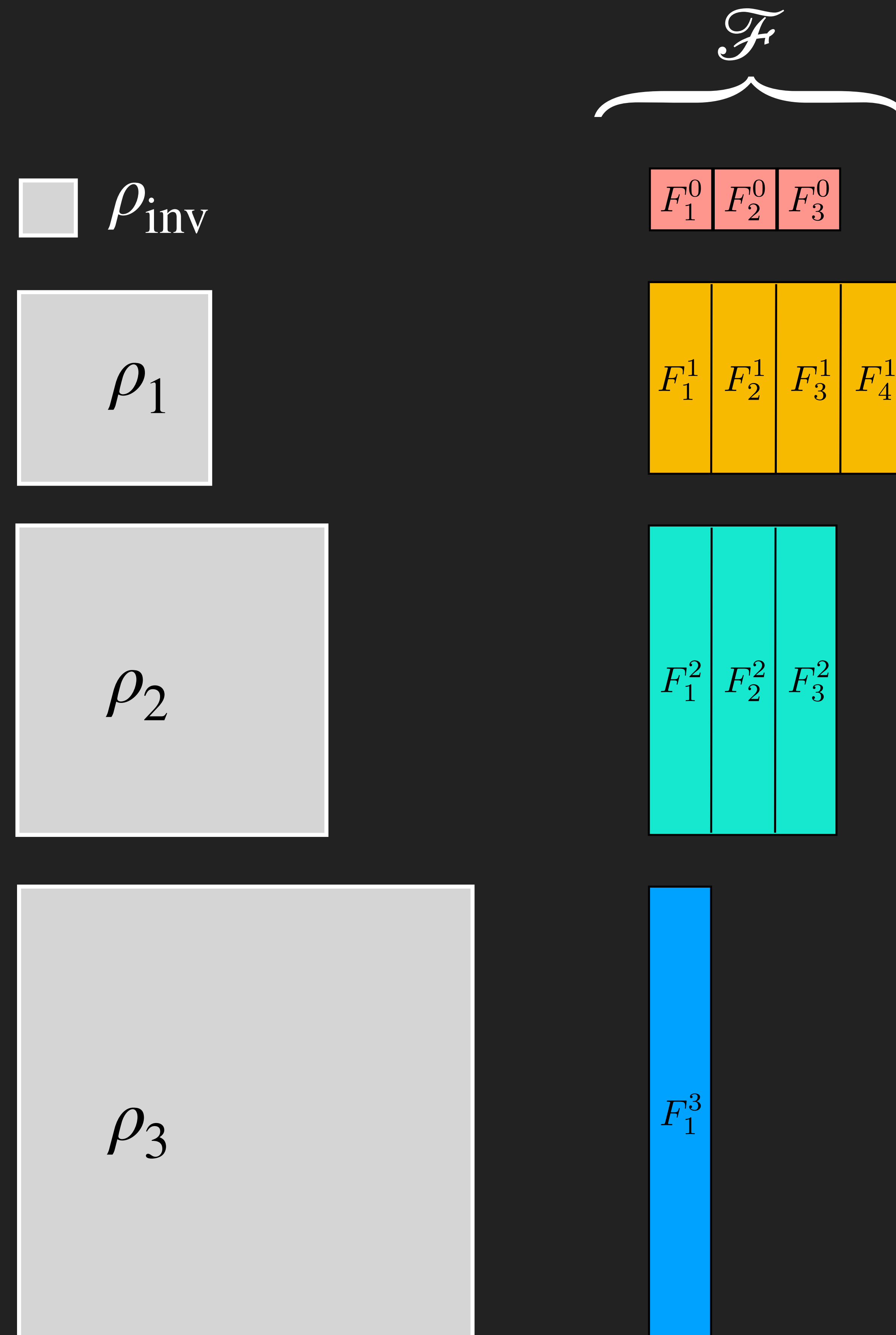
Finite-dimensional representations of the Lorentz group are **decomposable**, i.e. are direct sums of irreps.

$$W : R \to R', \quad W \cdot \rho(g) = \rho'(g) \cdot W$$

Schur's lemma implies that W acts as scalar multiplication on each irrep, and only linearly combines vectors of the same weight.

Activations need to be stored as collections of irreducible components.





II. CLEBSCH-GORDAN PRODUCT

The "only" bilinear equivariant operation mapping two representations to another one is the **tensor product**

$$\otimes : R_1 \times R_2 \to R_3$$

$$\rho_1(g) \otimes \rho_2(g) = C_{\rho_1,\rho_2} \left[\bigoplus_{\rho} \bigoplus_{1}^{\mu(\rho)} \rho(g) \right] C^{\dagger}_{\rho_1,\rho_2}$$

Since our linear operation requires knowledge of irreducible components, each product must be followed by a **Clebsch-Gordan decomposition**.

II. CLEBSCH-GORDAN PRODUCT

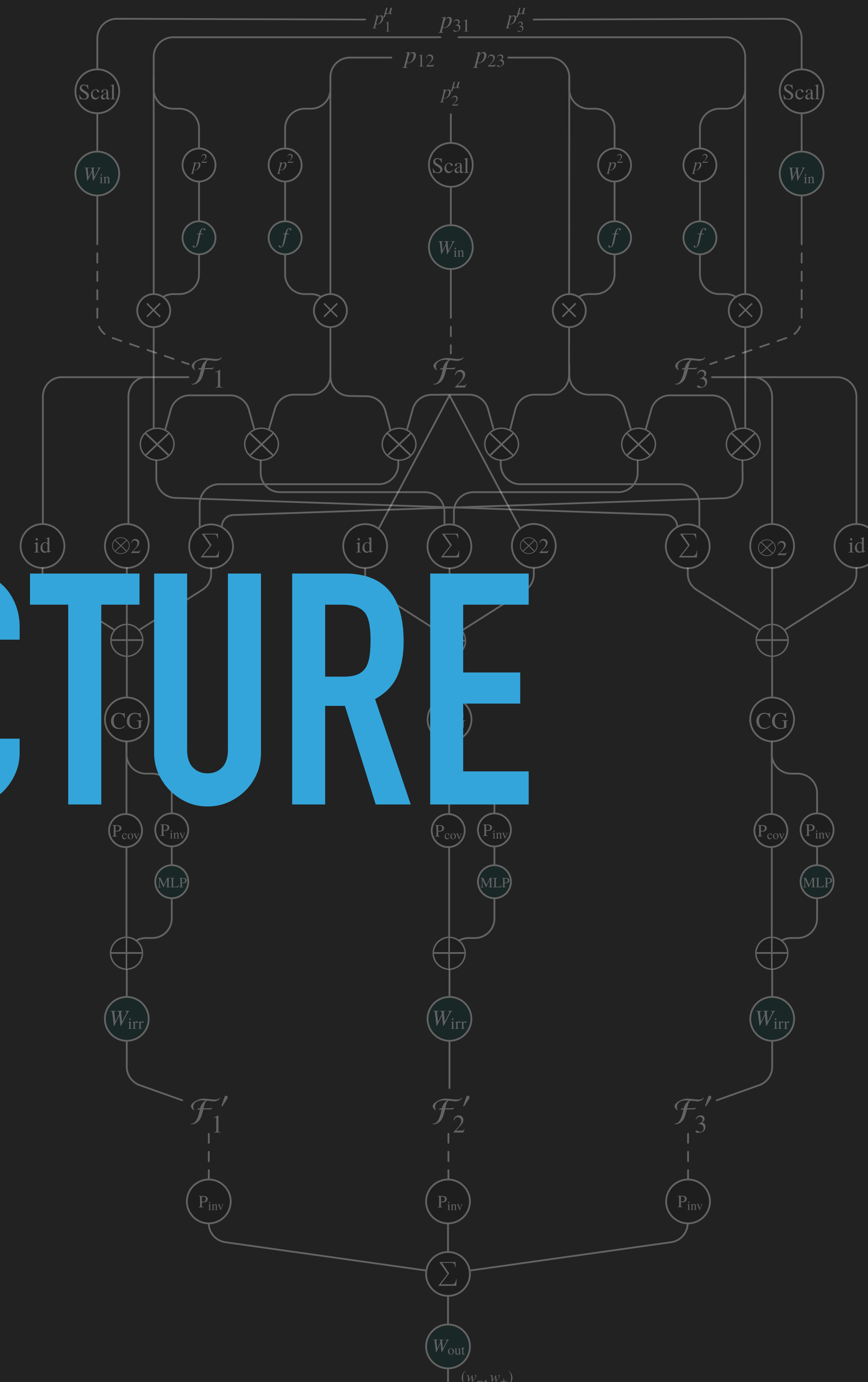
The "only" bilinear equivariant operation mapping two representations to another one is the **tensor product**

$$\otimes : R_1 \times R_2 \to R_3$$

$$\rho_1(g) \otimes \rho_2(g) = C_{\rho_1,\rho_2} \left[\bigoplus_{\rho} \bigoplus_{1}^{\mu(\rho)} \rho(g) \right] C^{\dagger}_{\rho_1,\rho_2}$$

Since our linear operation requires knowledge of irreducible components, each product must be followed by a **Clebsch-Gordan decomposition**.

ARCHITECTURE



- N particles with input 4-momenta p_i^μ
- N activations $\mathscr{F}_i$ at each level live in representations of the Lorentz group
- The update rule involves pair interactions

$$\mathscr{F}_i \mapsto W \cdot \left(\mathscr{F}_i \oplus \mathscr{F}_i^{\otimes 2} \oplus \sum_j f\left(p_{ij}^2\right) \cdot p_{ij} \otimes \mathscr{F}_j \right) \quad *$$

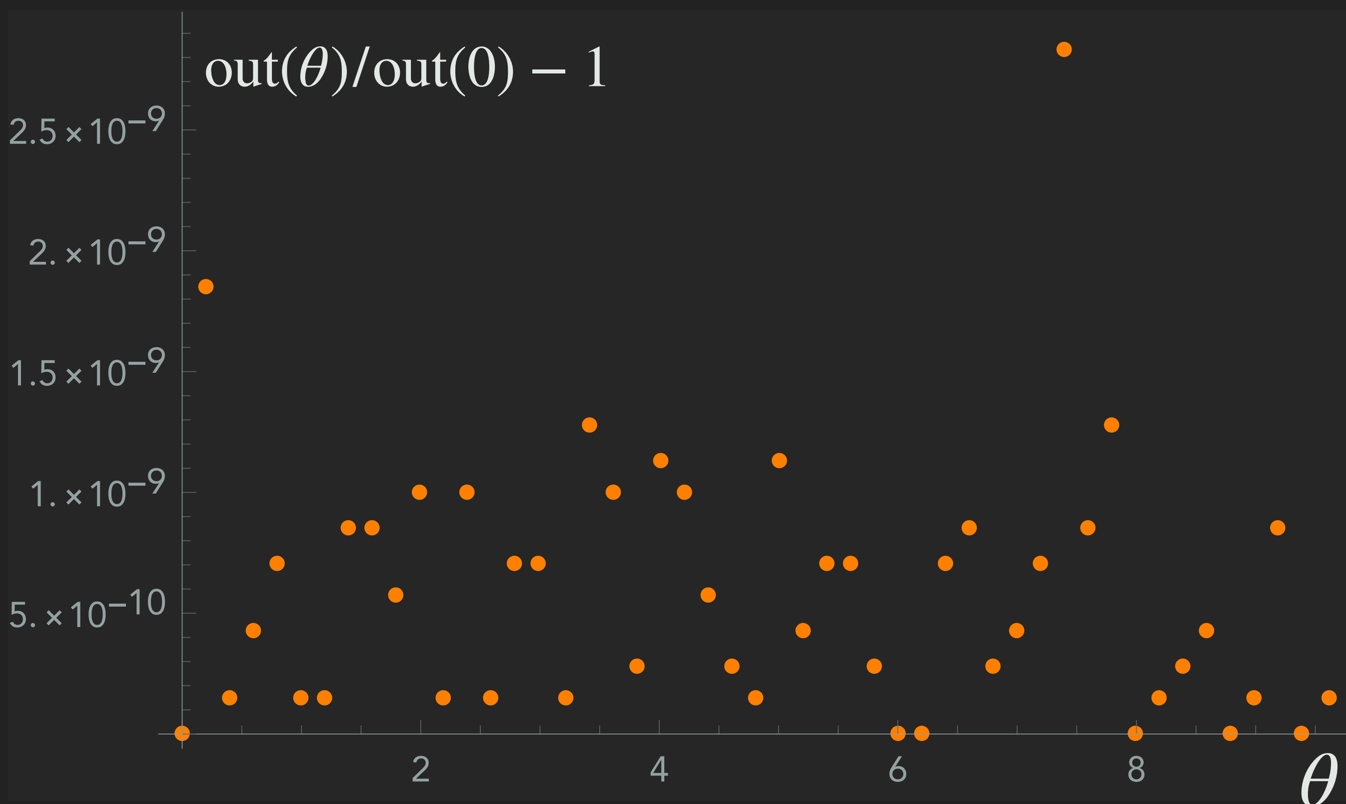
- Arbitrary traditional sub-networks can be applied to Lorentz invariants
- Output layer sums over i and projects onto invariants (or other irrep)

$$\mathscr{M}^2 = \frac{1}{4} \frac{(2g_c)^4}{((p_1 - p_3)^2 - m_\gamma c^2)^2} \left[p_1 \cdot p_3 + m_e c^2\right] \left[p_2 \cdot p_4 + m_\mu c^2\right]$$

IRC Safety to be addressed separately

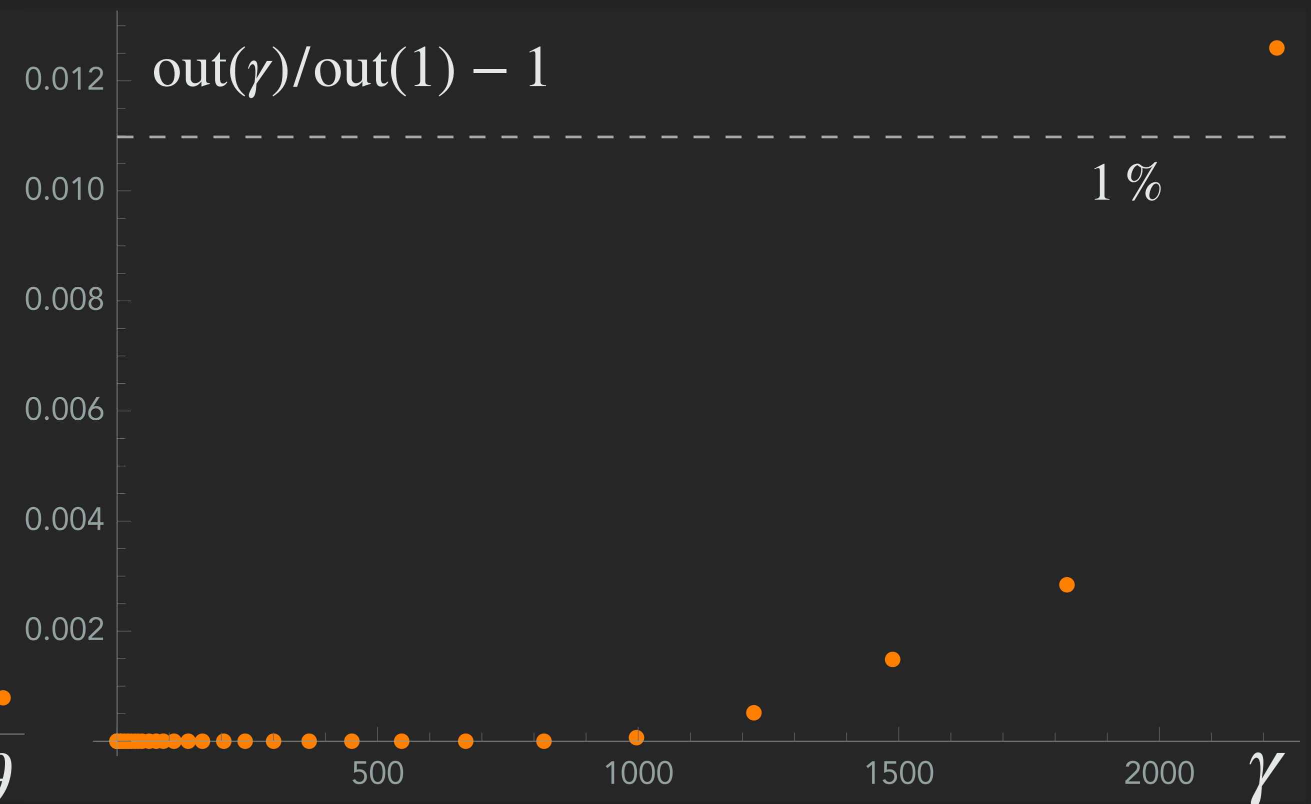
PERFORMANCE

Rotational invariance

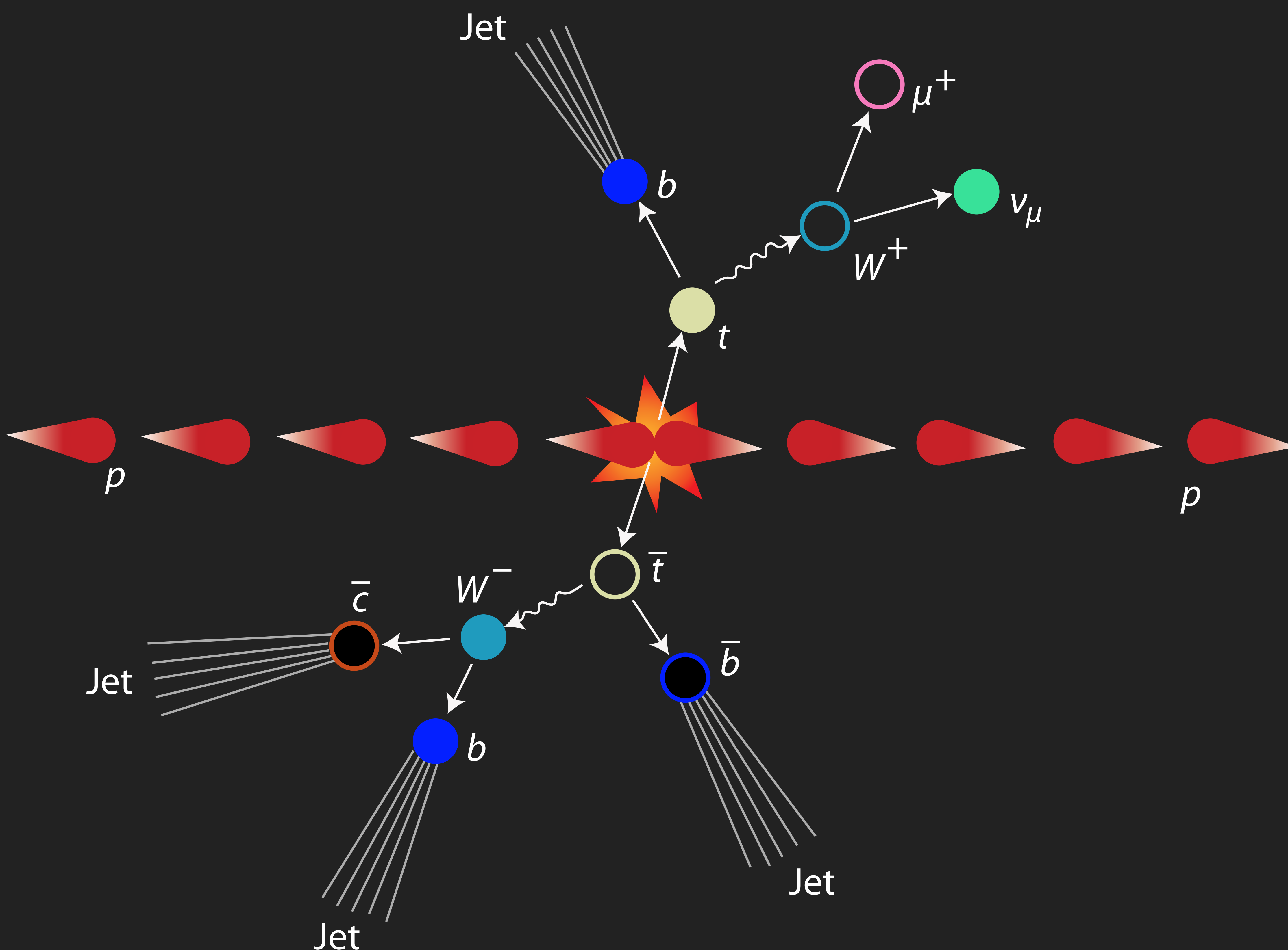


Relative invariance within 10^{-9} using double precision

Boost invariance

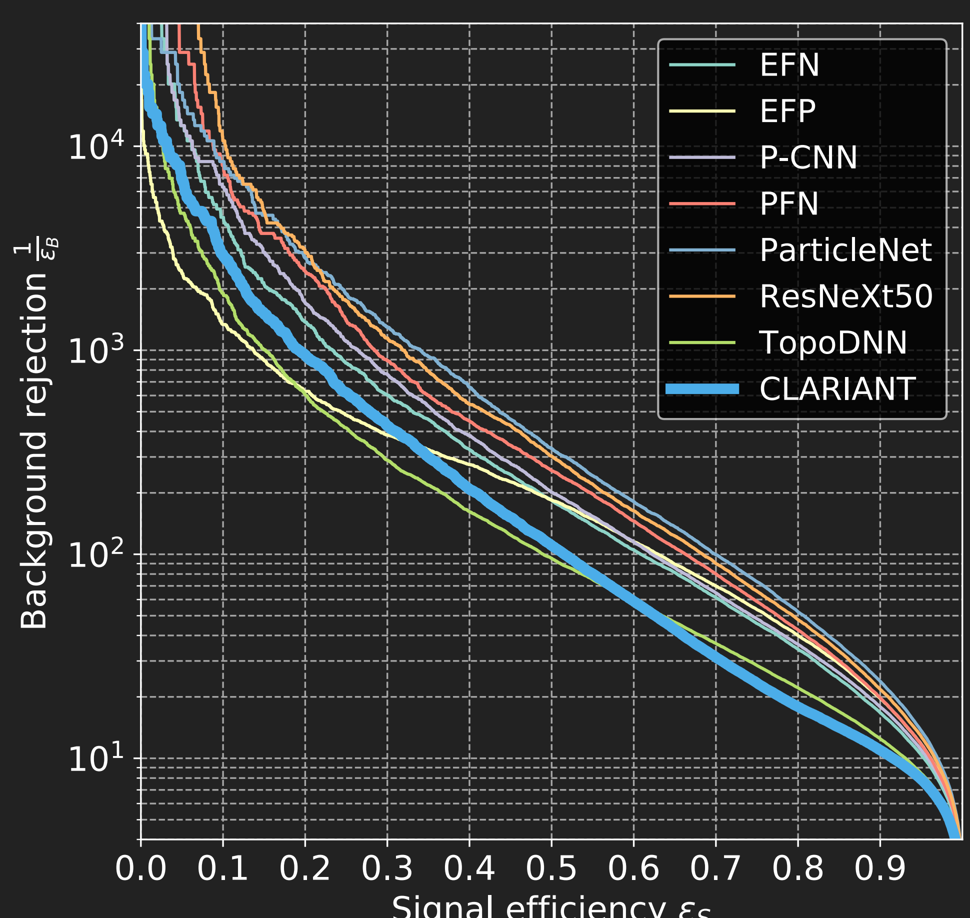


Relative invariance within 10^{-2} limited by floating point precision



Dataset: [Butter, Kasieczka, Russell & Russell (*Deep-learned Top Tagging with a Lorentz Layer*) arXiv: 1707.08966]

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN	0.981	0.931	836±17	859±67	966±20	705k
LoLa	0.980	0.929	722±17	768±11	765±11	127k
LDA	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials	0.980	0.932	384			1k
Energy Flow Network	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network	0.982	0.932	891±18	1063±21	1052±29	82k
CLARIANT	0.970	0.922	426			3k



[Kasieczka et. al. (ML Landscape of top taggers, 2019)]

[Xia et. al. (ResNeXt, 2016)]
[Moore at. al. (Multi-body N-subjettiness, 2018)]
[Qu & Gouskos (ParticleNet, 2019)]
[Thaler et. al. (Energy/Particle Flow Polynomials, 2018/19)]
[Butter et. al. (LoLa – Lorentz Layer, 2018)]

FUTURE WORK AND POTENTIAL EXTENSIONS

- Complete hyperparameter optimization;
- Include particle information (label, charge, spin, …);
- Regression tasks and measurements:
  - invariant mass detection,
  - covariant 4-momentum measurements;
- Detection of hidden symmetries;
- Multiple symmetries combined (Standard Model?);

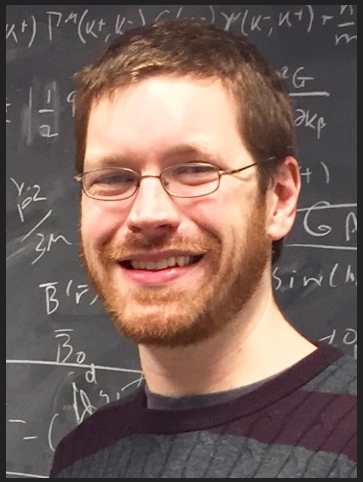
THE UNIVERSITY OF CHICAGO
Physics Department
Computer Science Department

ENRICO FERMI INSTITUTE

FLATIRON INSTITUTE

Center for Data and Computing
AT THE UNIVERSITY OF CHICAGO

KADANOFF CENTER
FOR THEORETICAL PHYSICS


Brandon Anderson

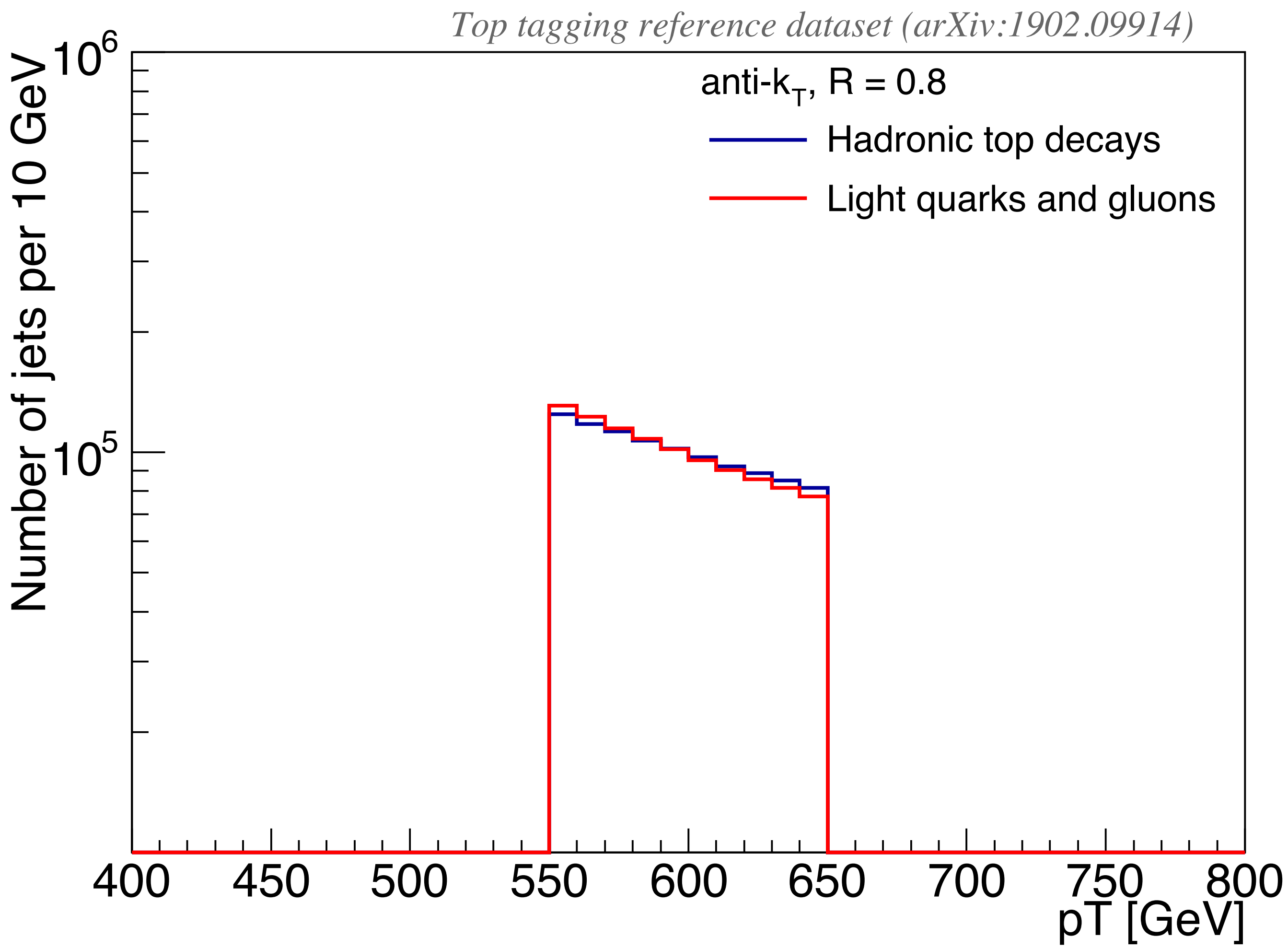

Risi Kondor

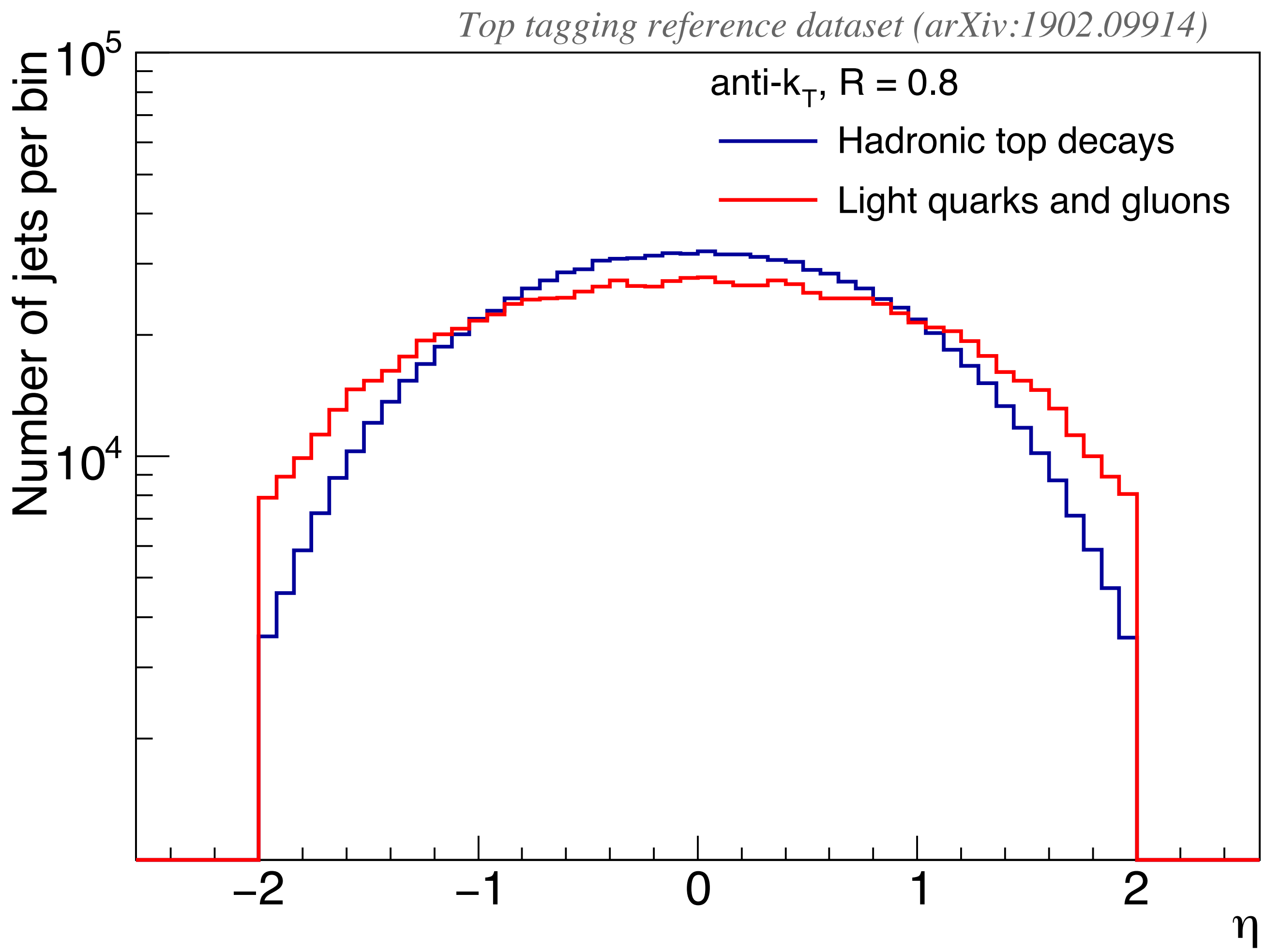

David Miller

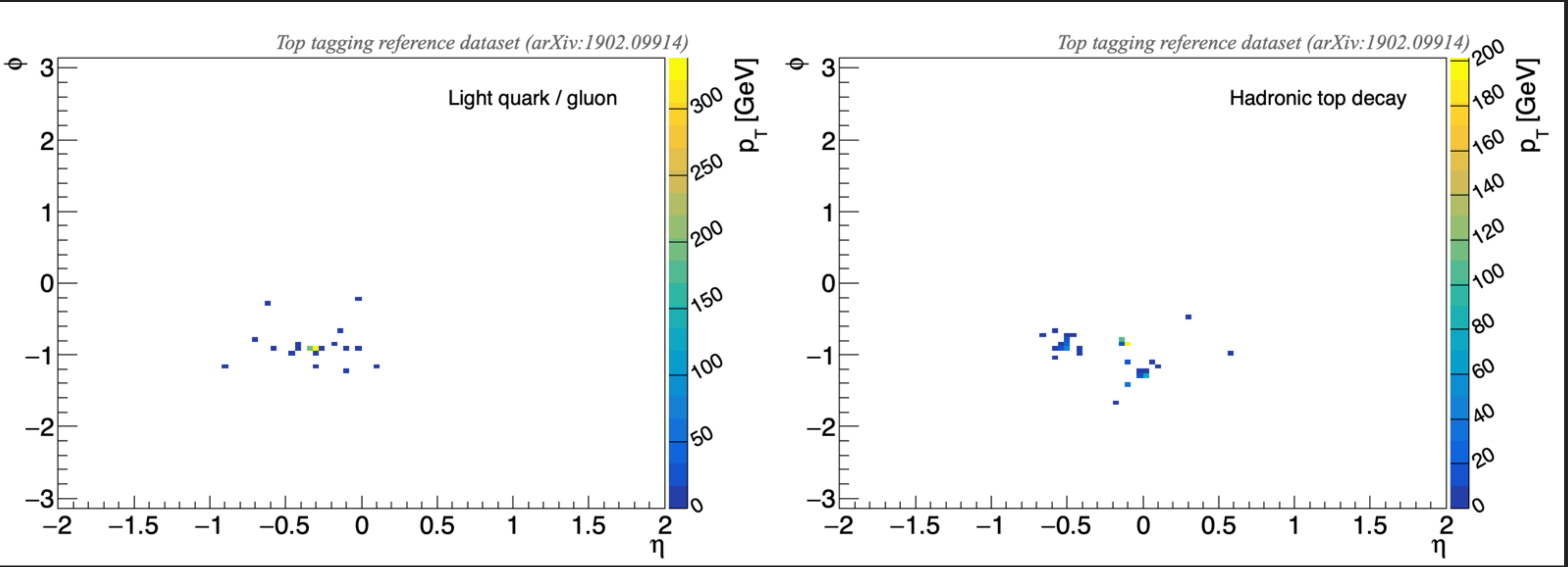

Jan Offermann

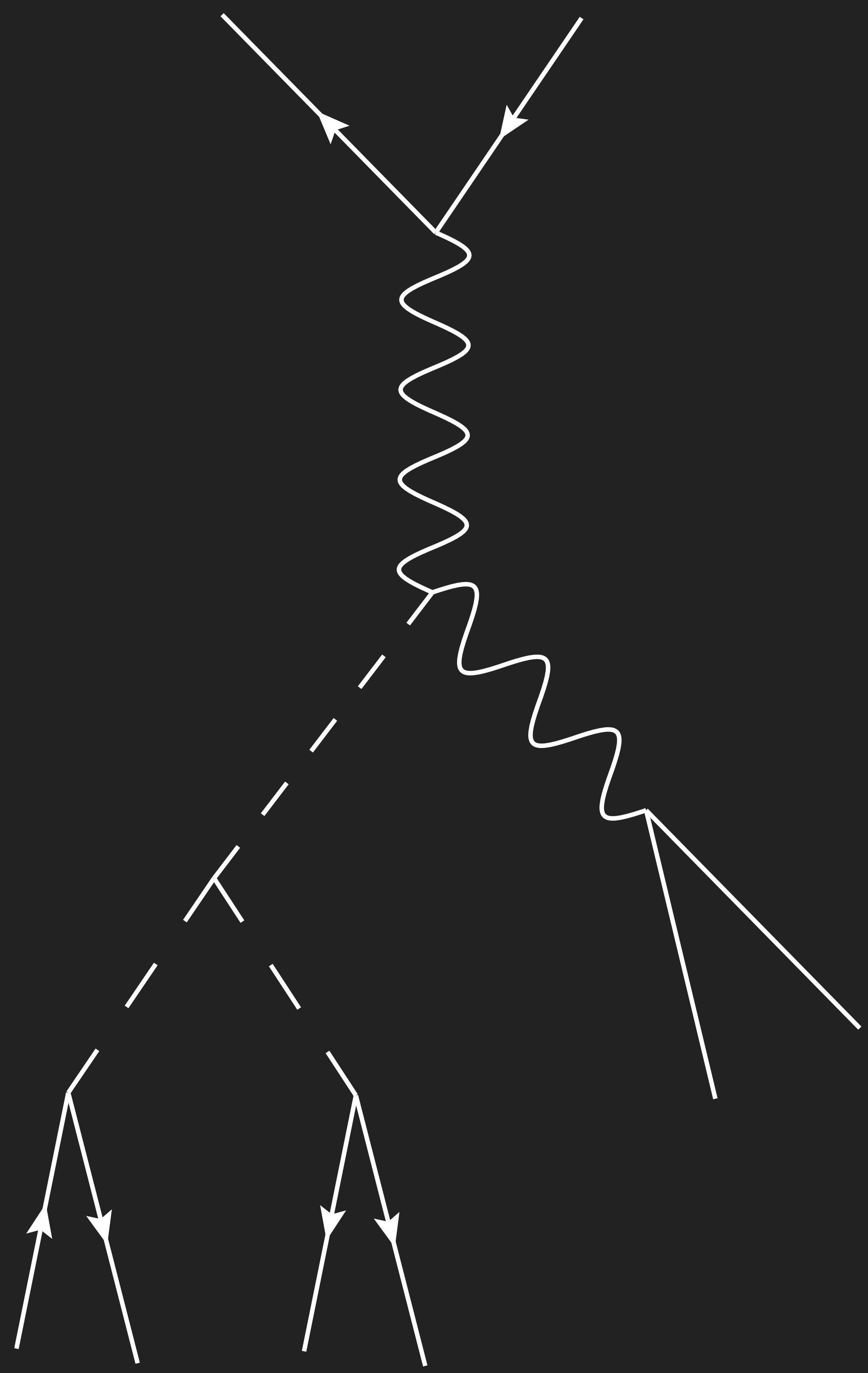

Marwah Roussi

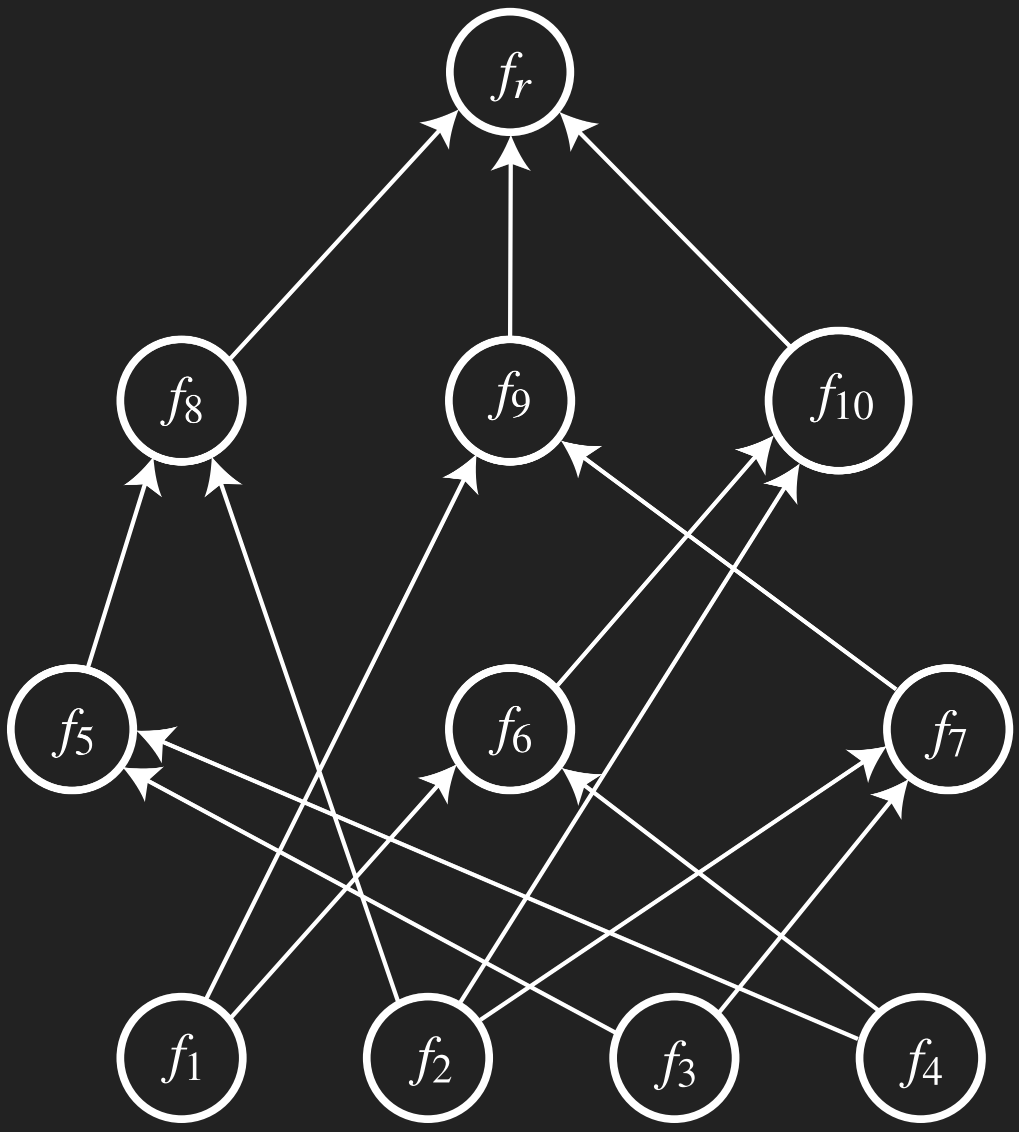
DATASET ENERGY DISTRIBUTIONS



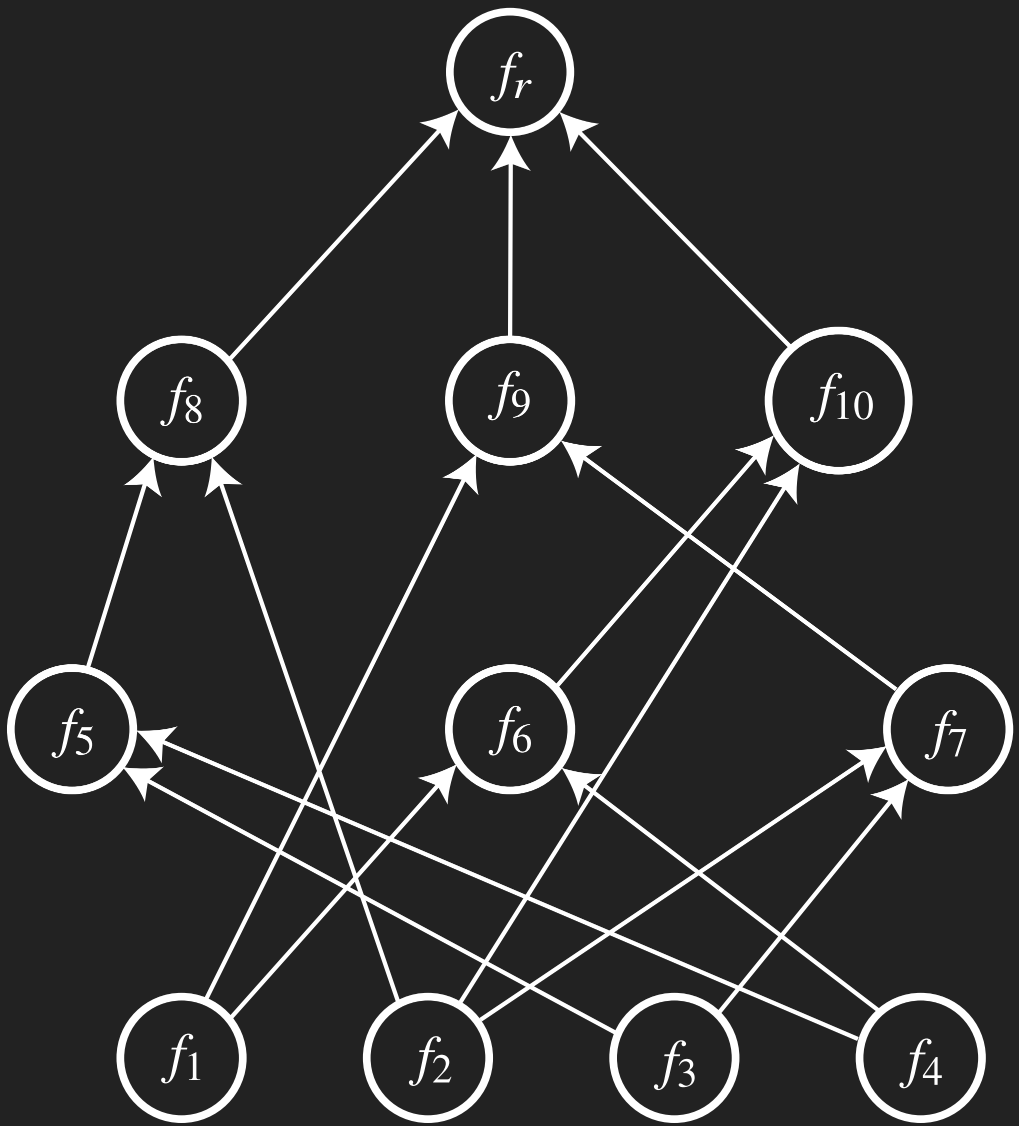



Top tagging reference dataset (arXiv:1902.09914)
Light quark / gluon
φ
η
p_T [GeV]
Top tagging reference dataset (arXiv:1902.09914)
Hadronic top decay
φ
η
p_T [GeV]

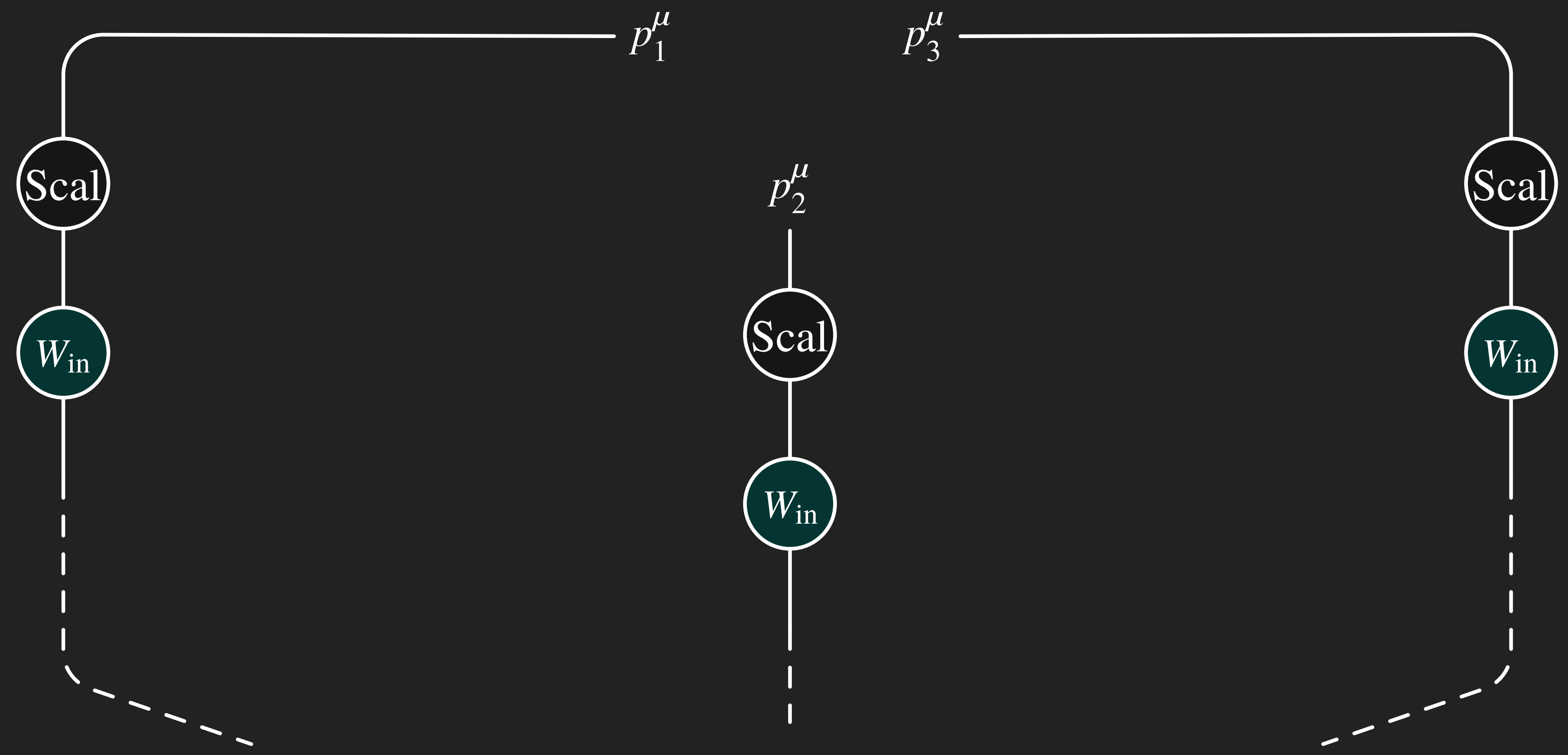



f_r
f_8
f_9
f_{10}
f_5
f_6
f_7
f_1
f_2
f_3
f_4

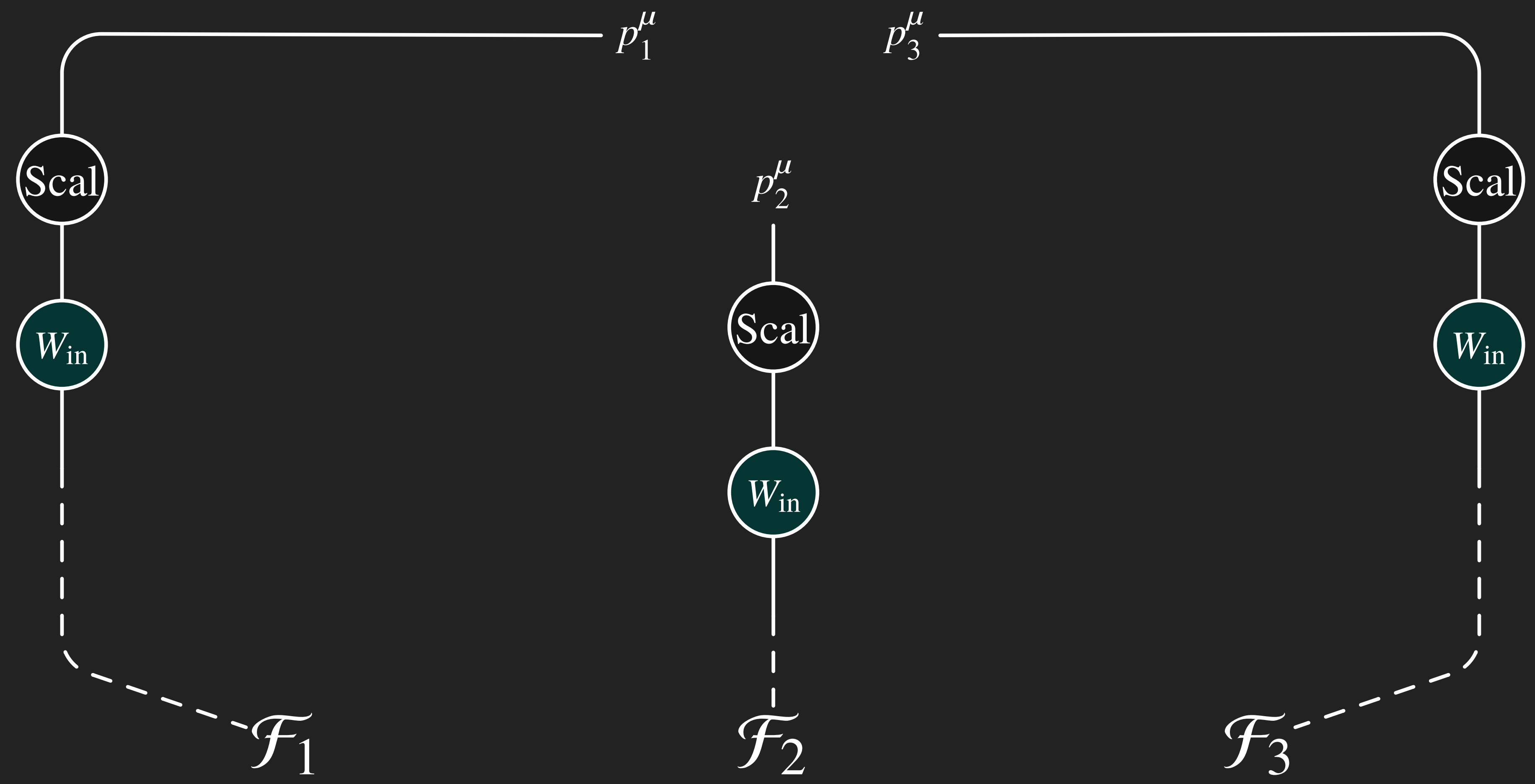
ATLAS



INPUT LAYER

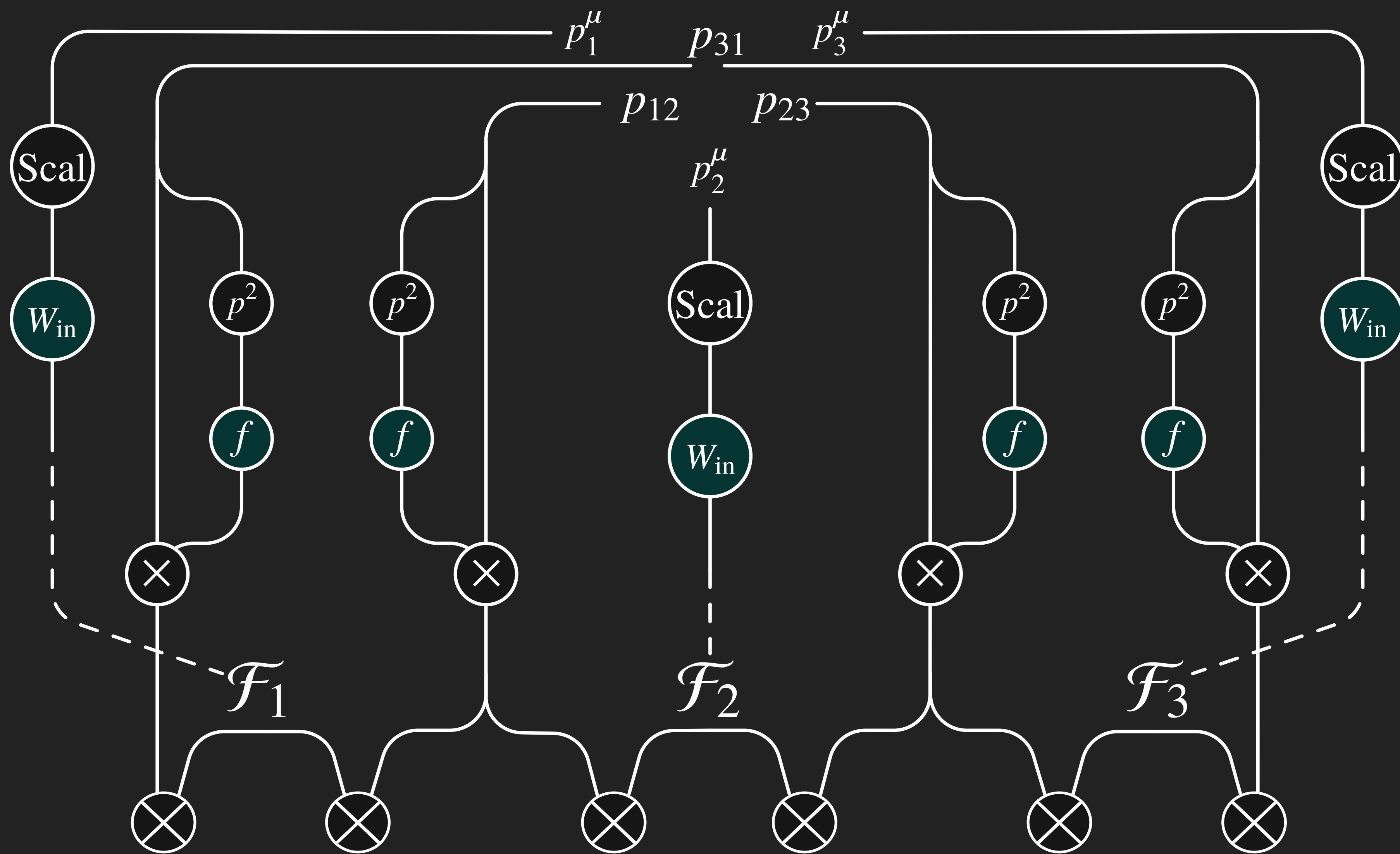


INPUT LAYER



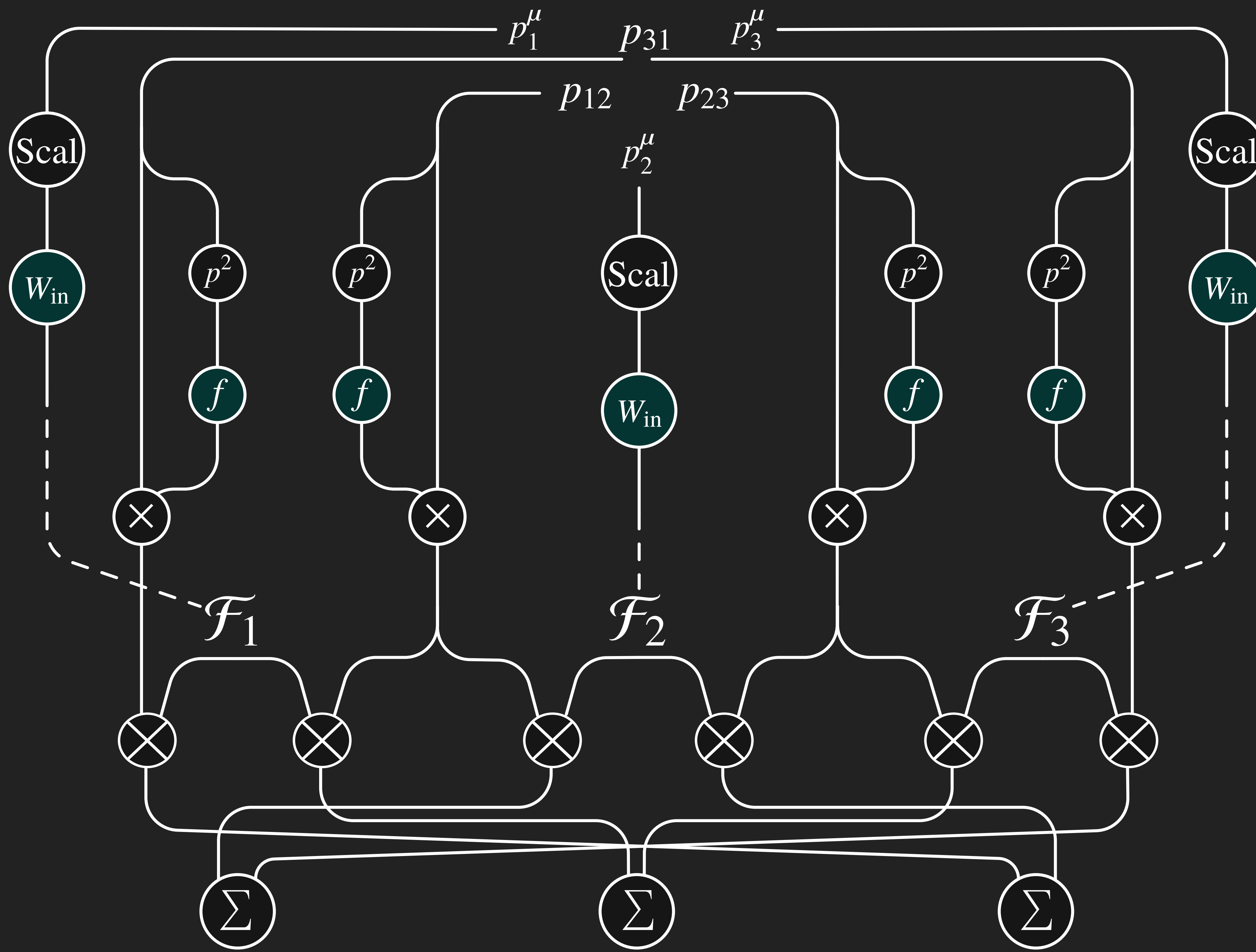
p_1^μ
p_3^μ
p_2^μ
Scal
Scal
Scal
W_{in}
W_{in}
W_{in}
$\mathcal{F}_1$
$\mathcal{F}_2$
$\mathcal{F}_3$

INPUT LAYER



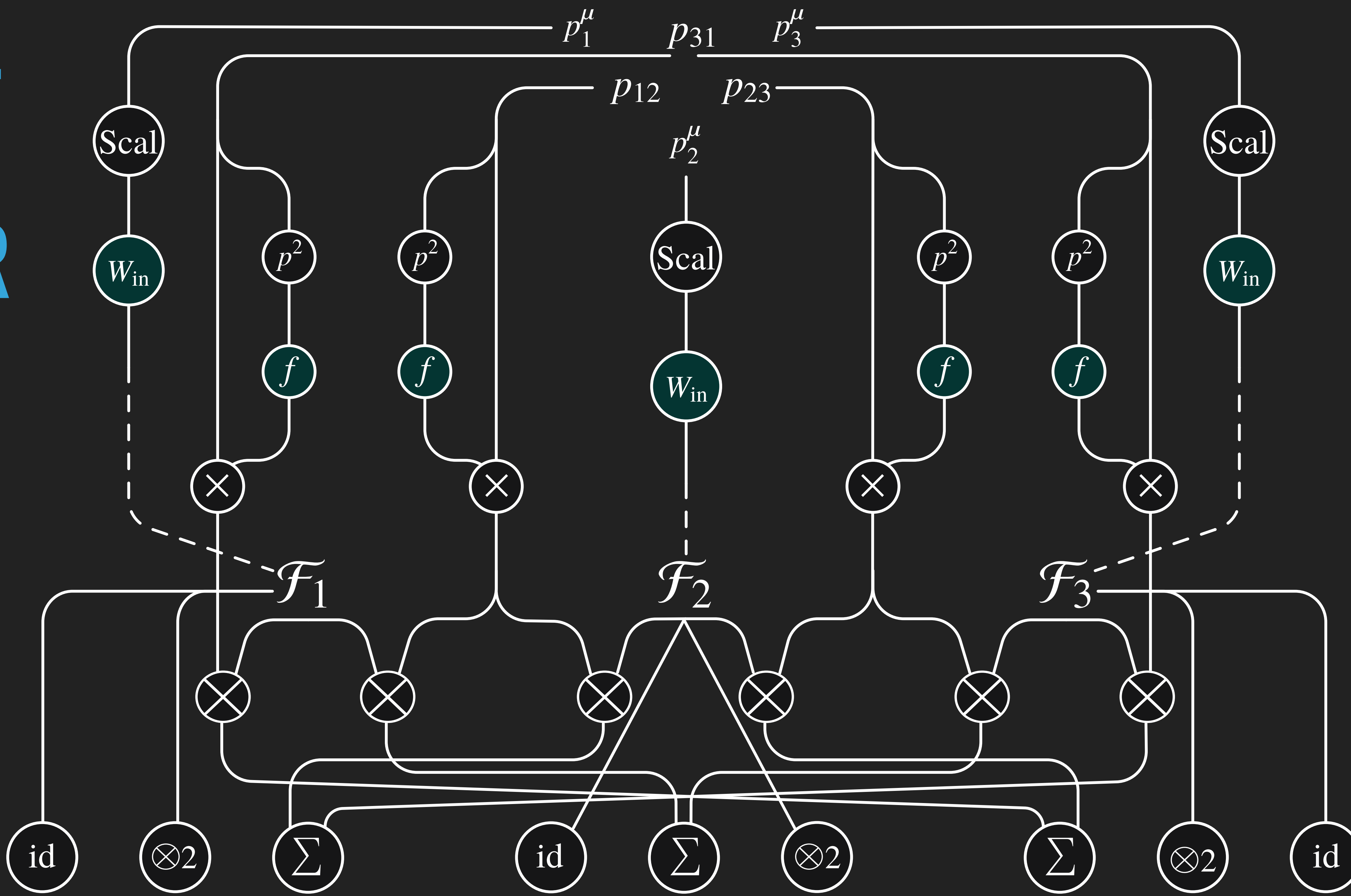
ITERATED CG LAYER

INPUT LAYER



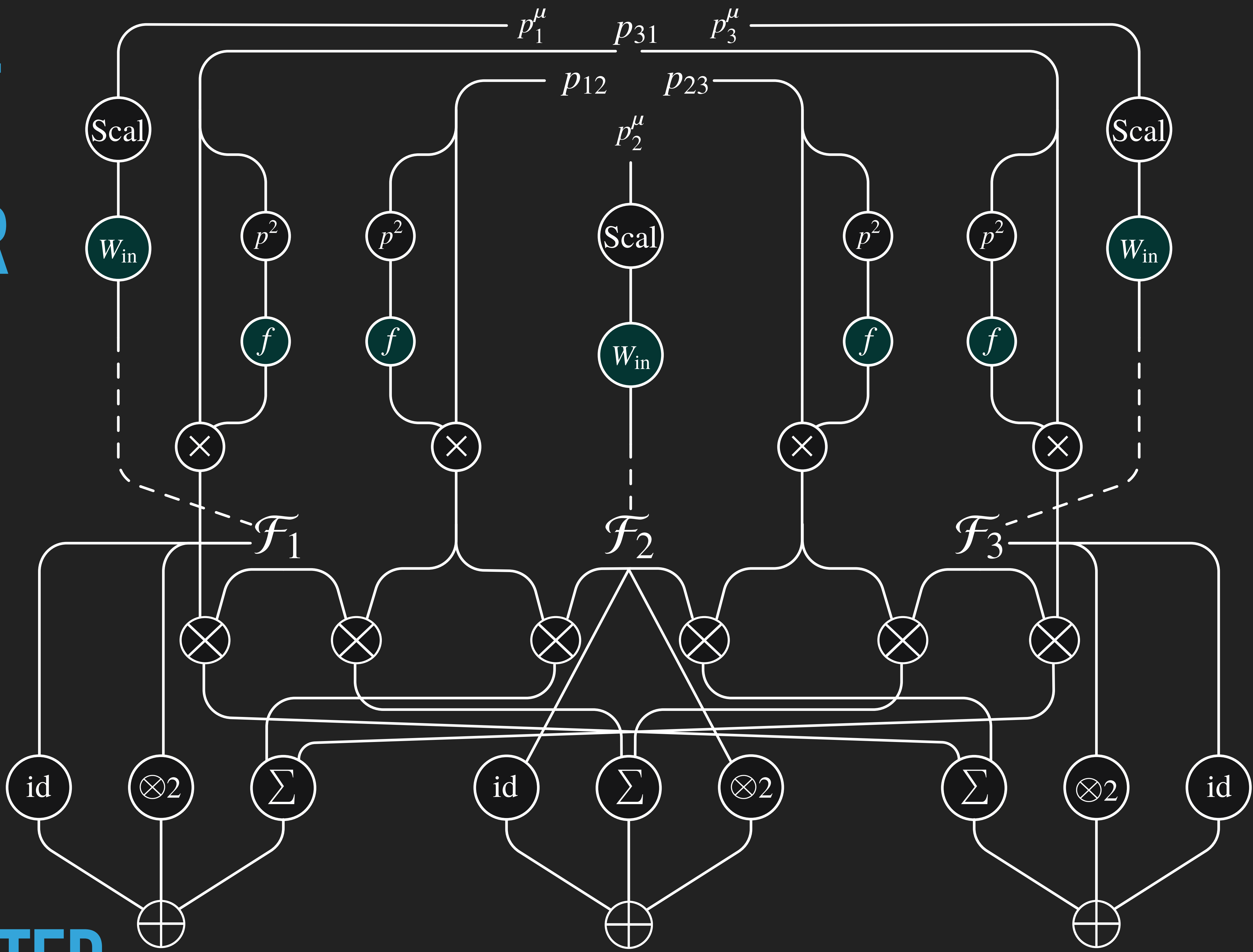
ITERATED CG LAYER

INPUT LAYER

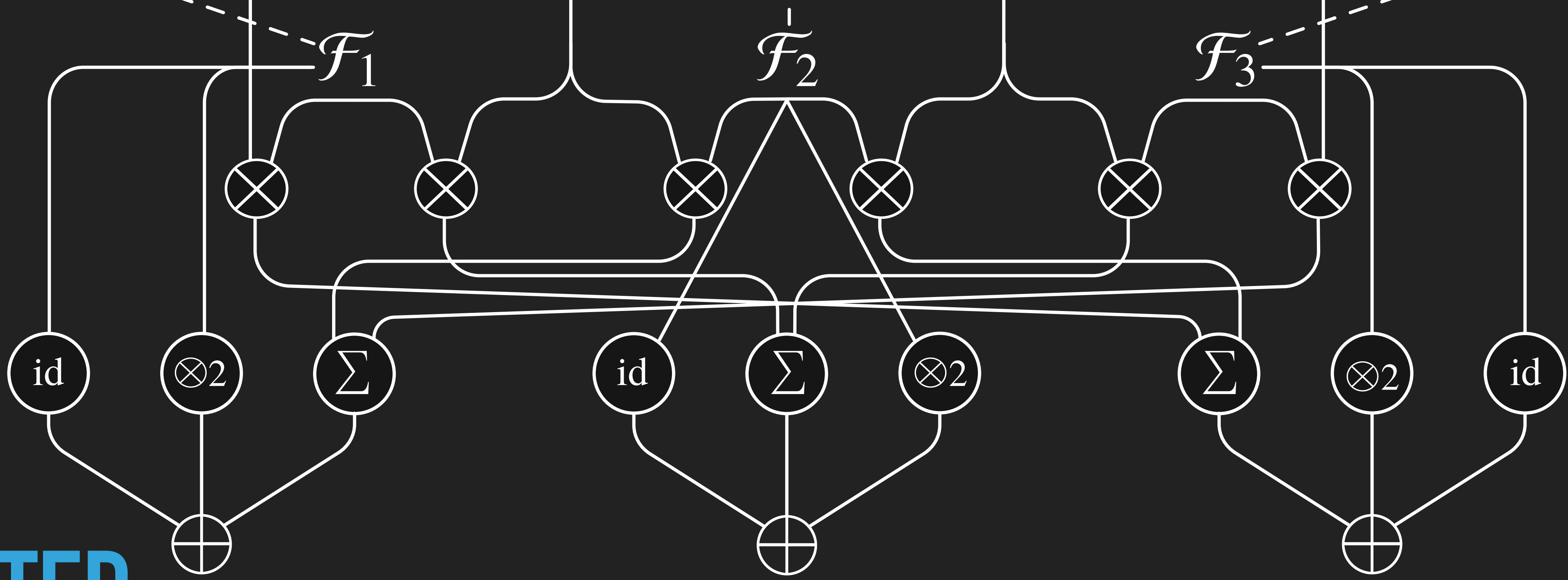


ITERATED CG LAYER

INPUT LAYER

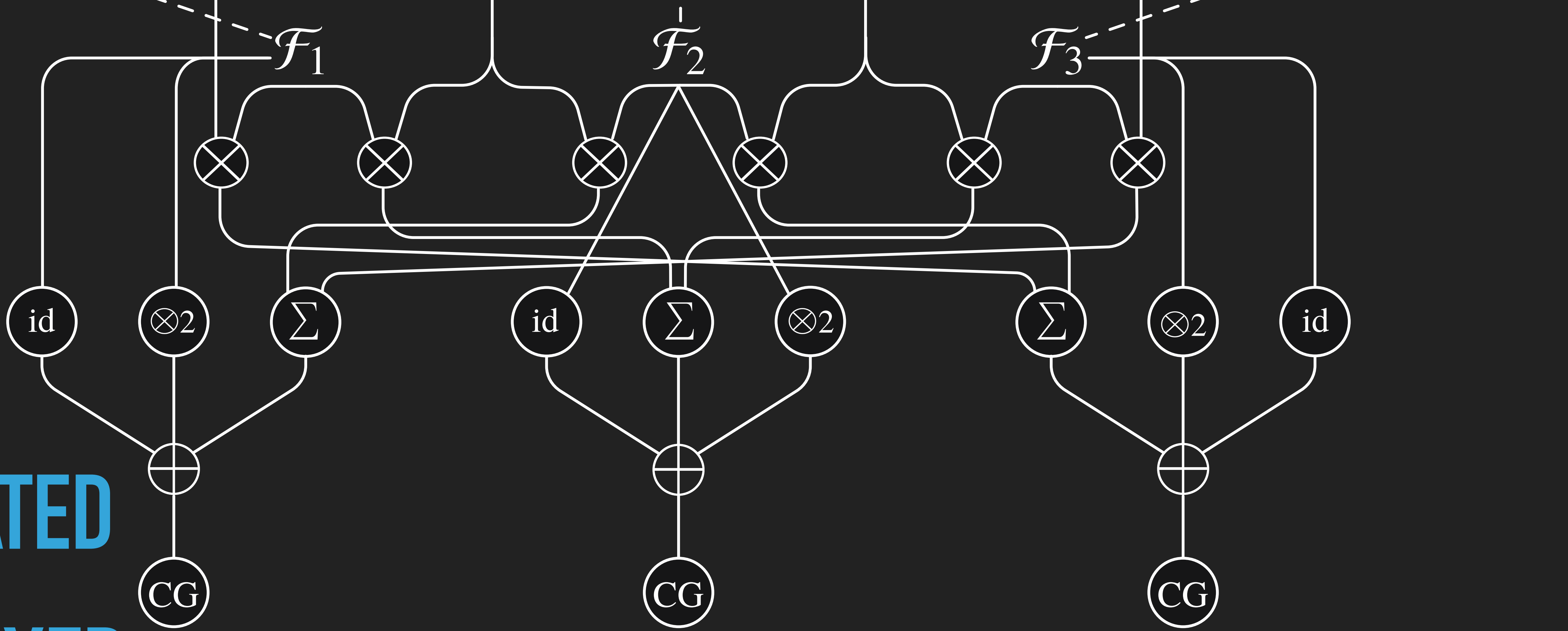


ITERATED CG LAYER

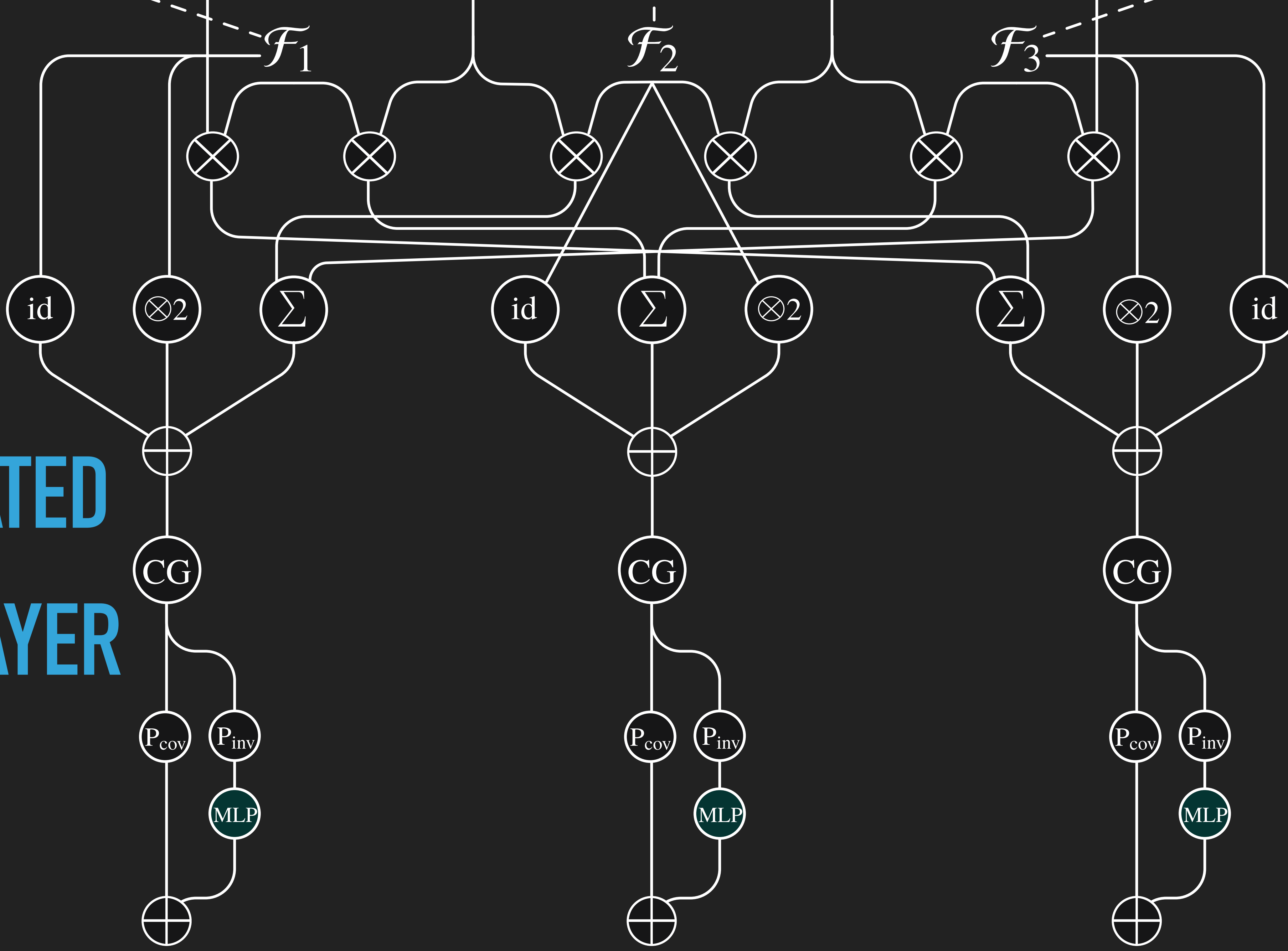


ITERATED CG LAYER

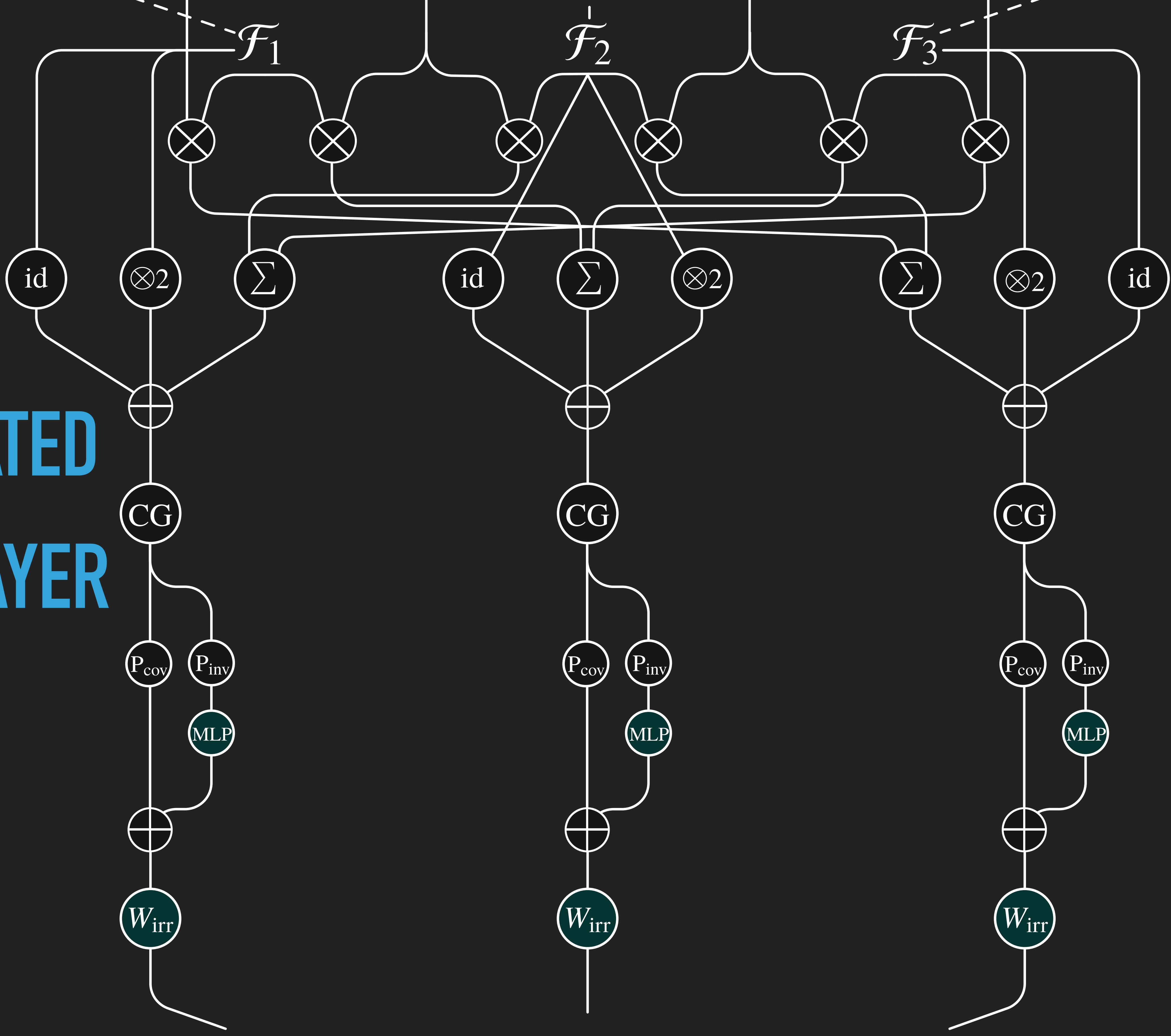
ITERATED CG LAYER



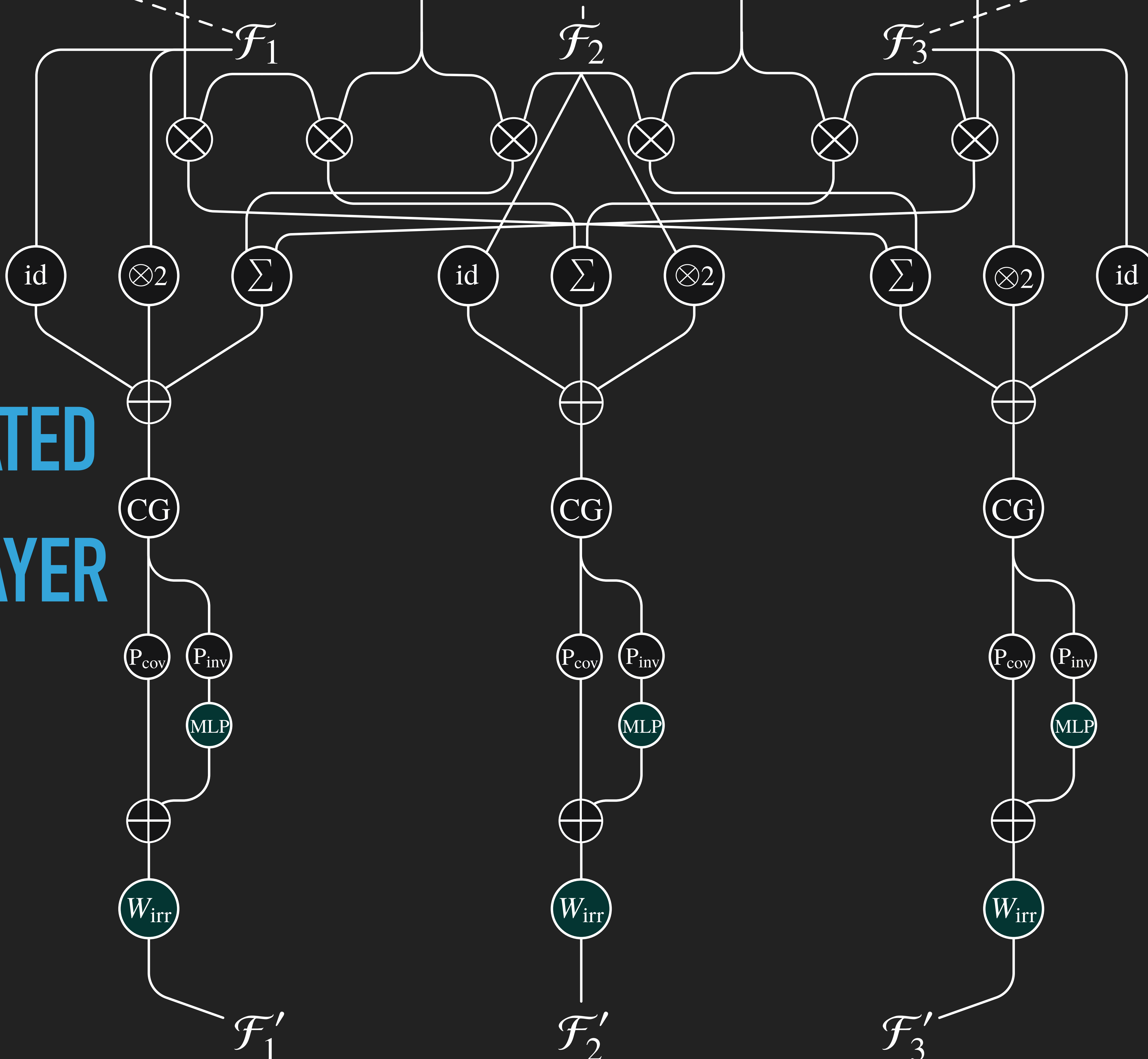
ITERATED CG LAYER



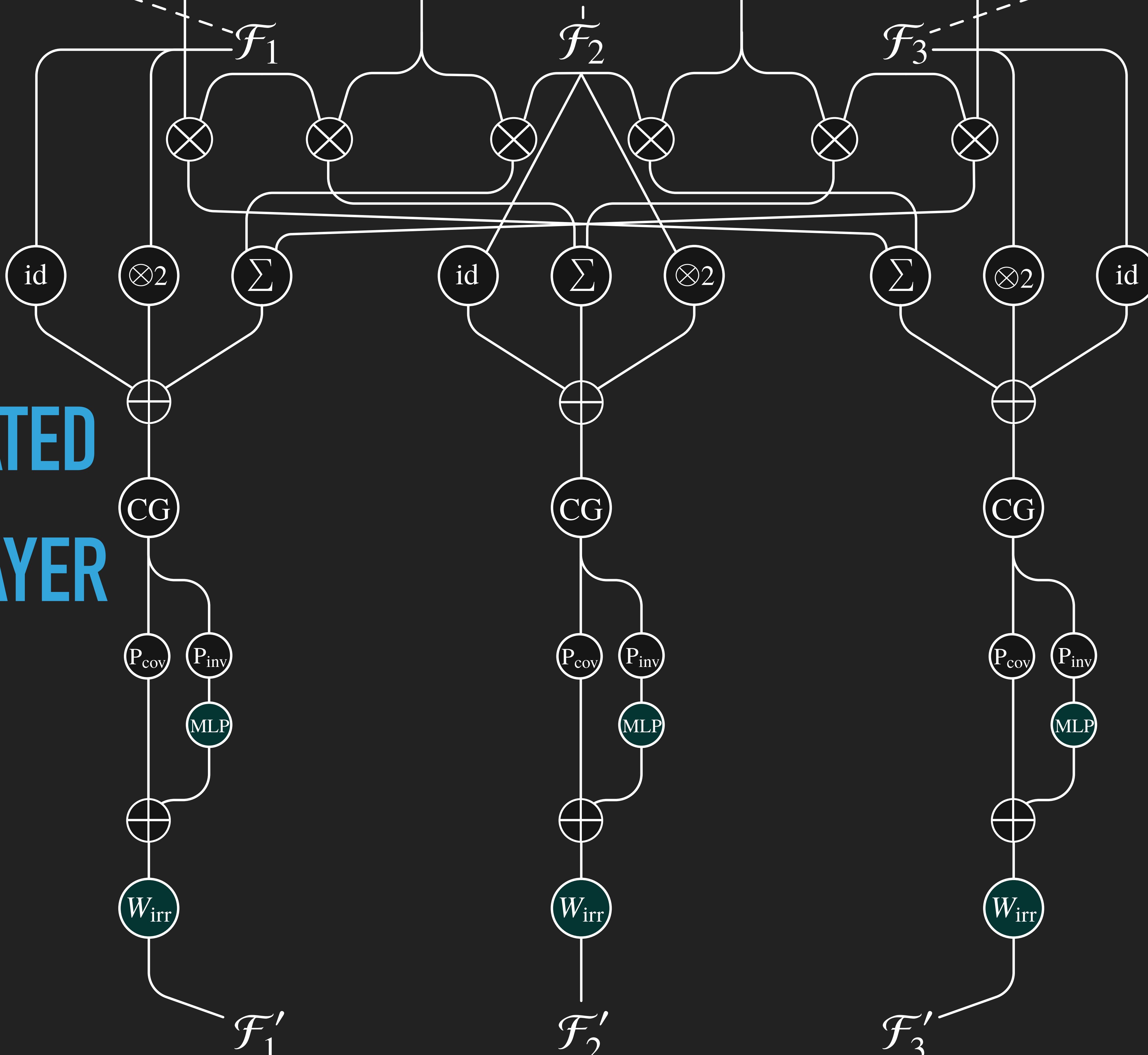
ITERATED CG LAYER



ITERATED CG LAYER

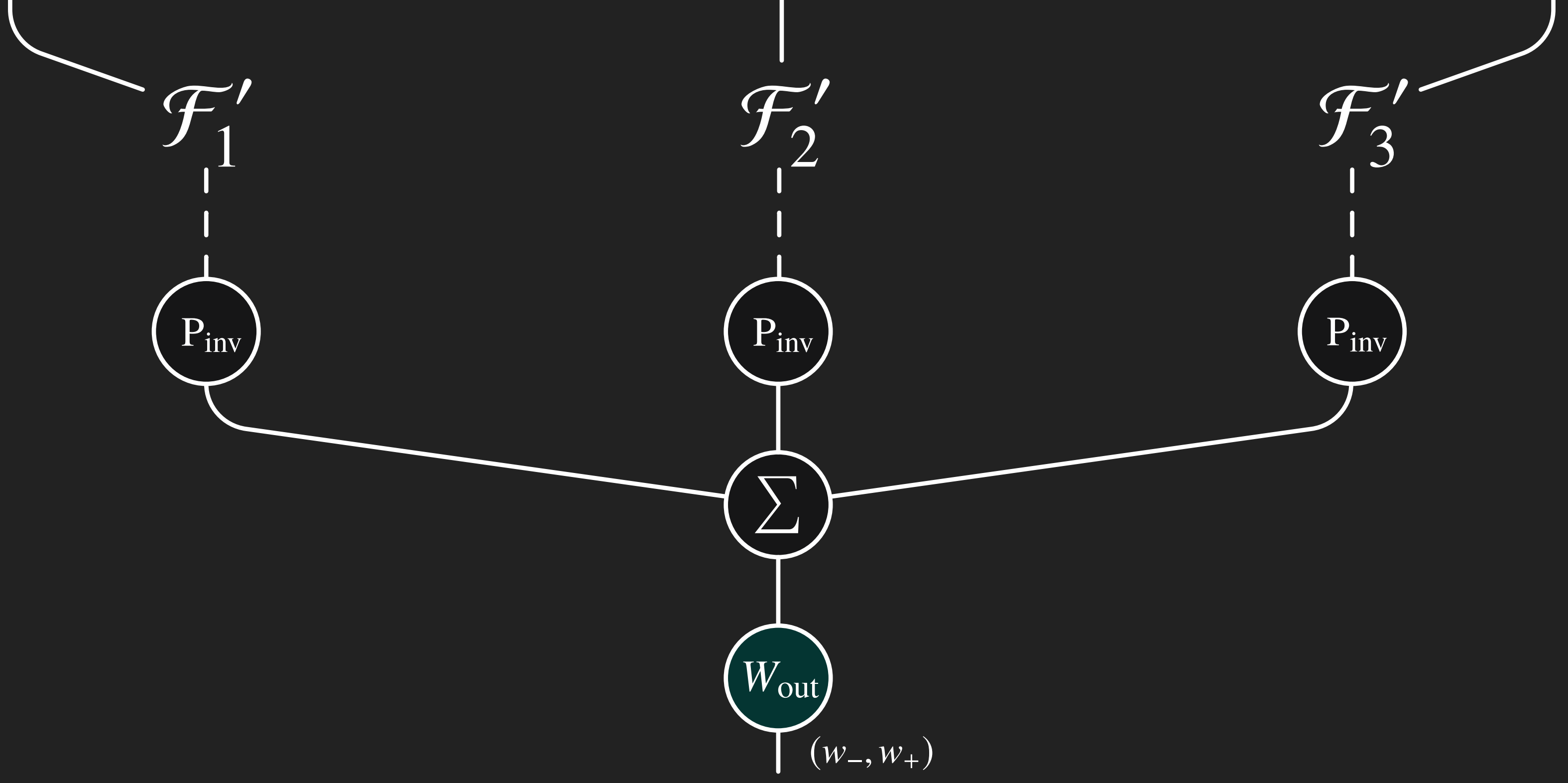


ITERATED CG LAYER

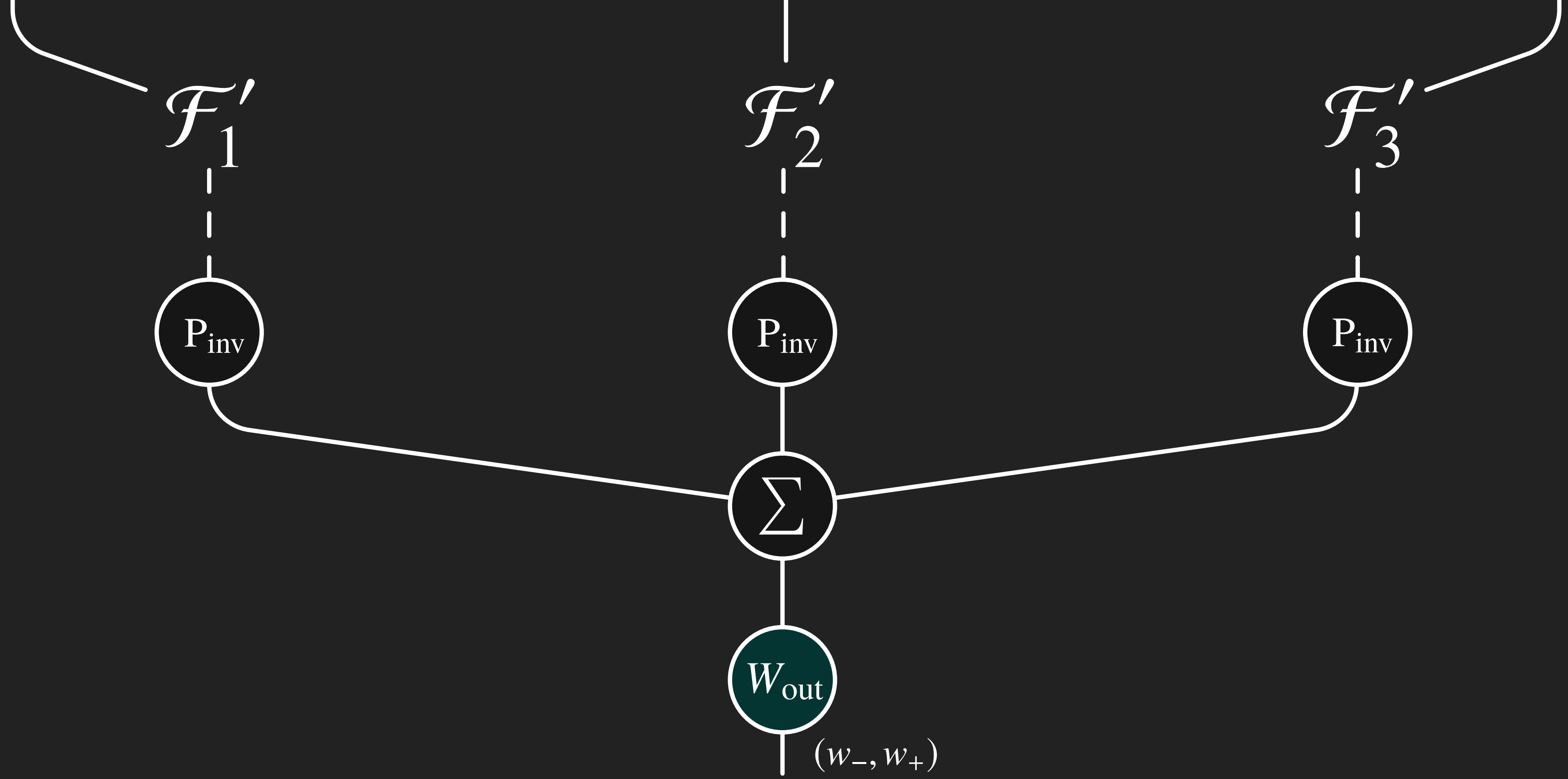



$\mathcal{F}_1'$
$\mathcal{F}_2'$
$\mathcal{F}_3'$

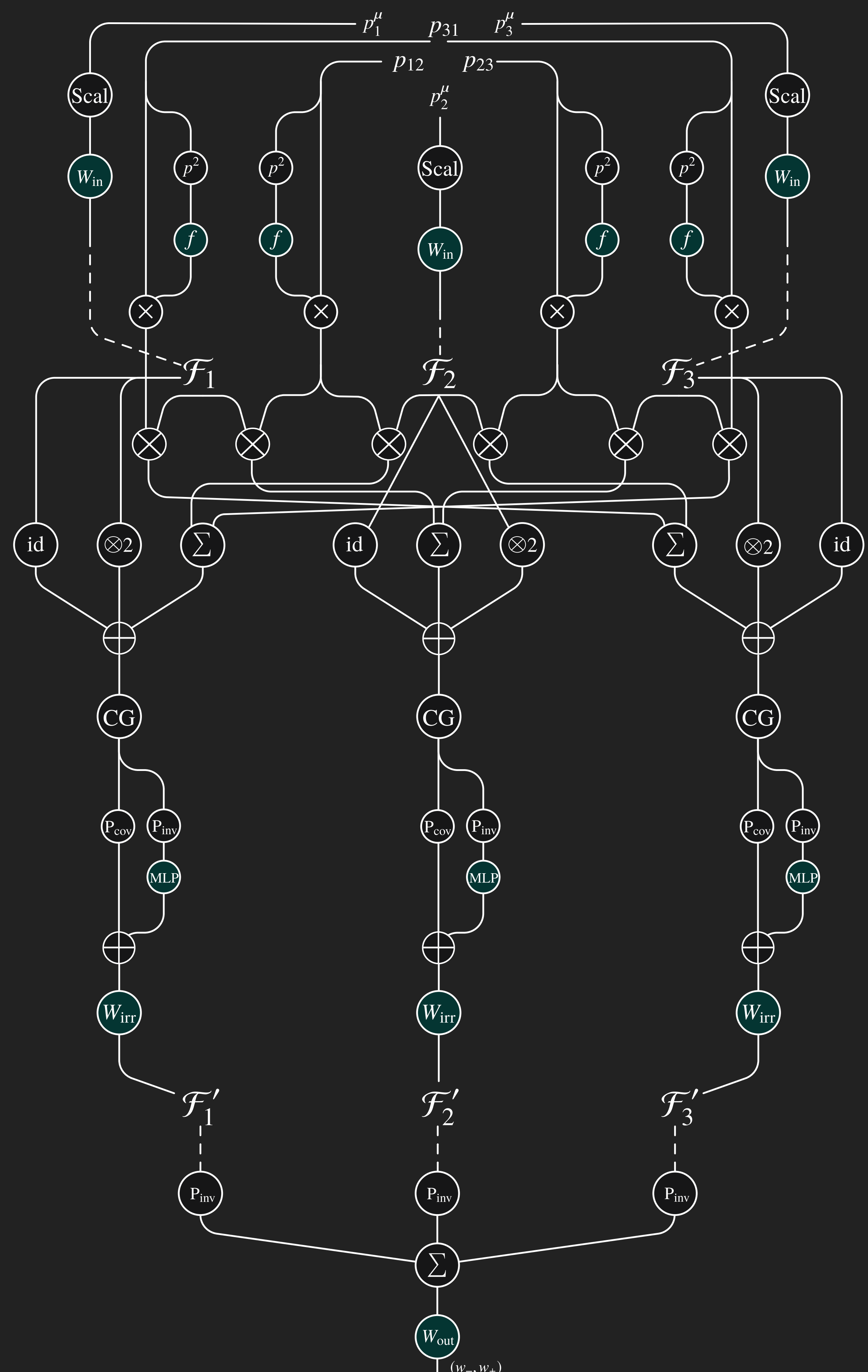
OUTPUT
LAYER



OUTPUT LAYER



CLARIANT



p_1^μ
p_{31}
p_3^μ
p_{12}
p_{23}
p_2^μ
Scal
W_{in}
p^2
f
$\mathcal{F}_1$
$\mathcal{F}_2$
$\mathcal{F}_3$
id
$\otimes 2$
$\sum$
CG
P_{cov}
P_{inv}
MLP
W_{irr}
$\mathcal{F}_1'$
$\mathcal{F}_2'$
$\mathcal{F}_3'$
W_{out}
(w_-, w_+)