



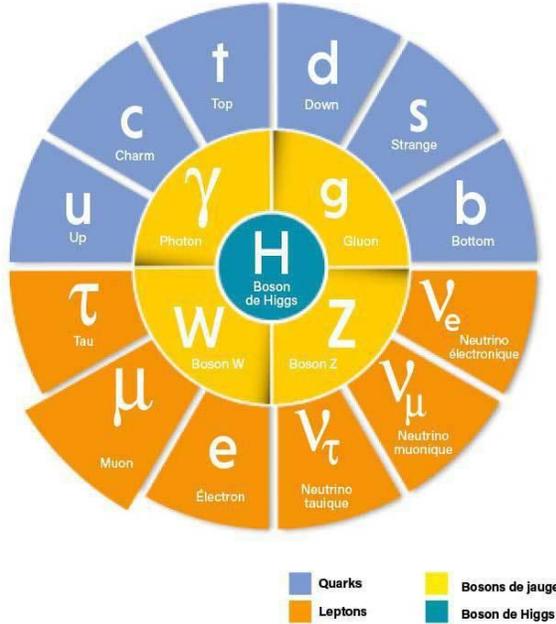
Trouver le méson B d'origine à l'aide d'un réseau de neurones à graphes à Belle II

BROUILLARD Alizée
Superviseur: DUJANY Giulio
24/06/2021

Sommaire

 Contexte scientifique, expérience Belle II3
 Cadre du projet & réseau de neurones à graphes6
 Données utilisées21
 Résultats & discussion22
 Conclusion & perspectives30
 Bibliographie.....33

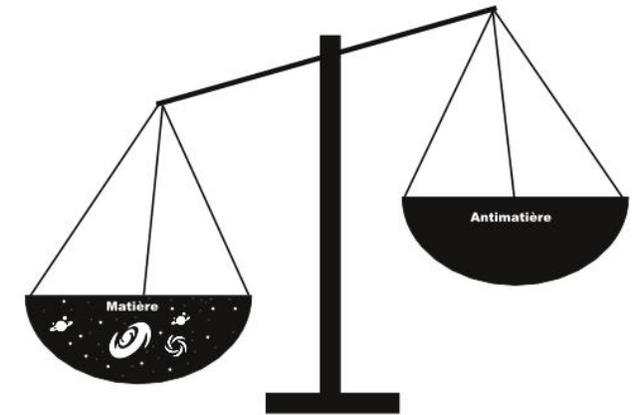
Contexte scientifique



➤ Modèle standard de la physique des particules

➤ Problèmes :

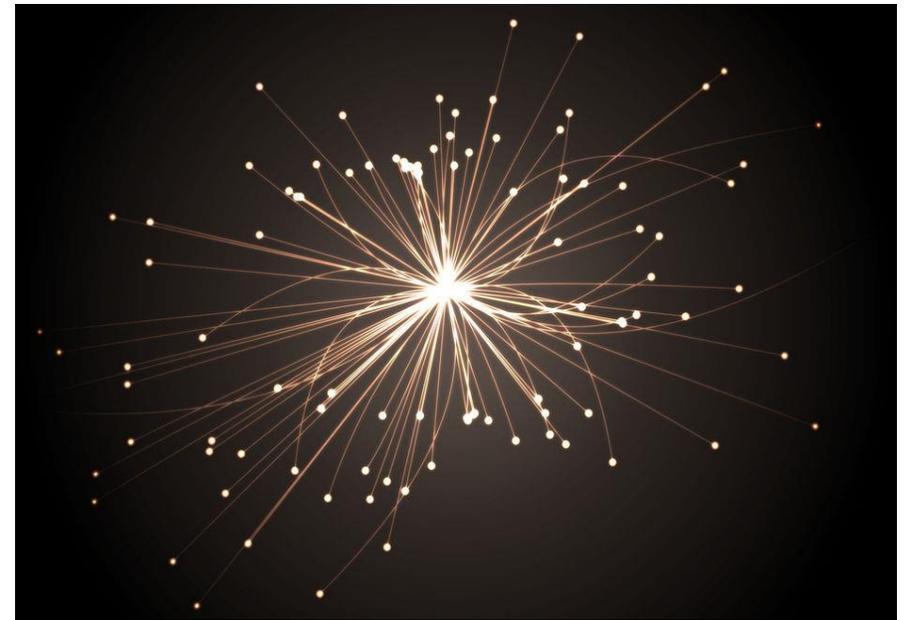
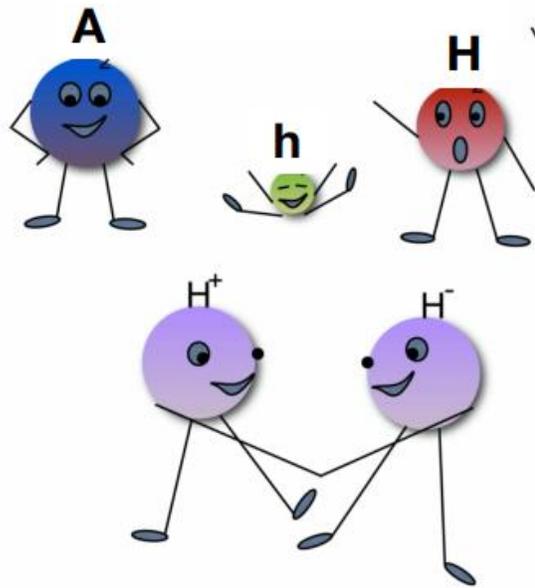
- gravitation
- asymétrie entre matière et anti-matière
- existence de la matière noire
- etc.



➤ Recherche d'une nouvelle physique

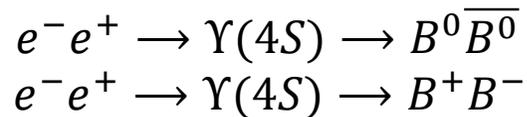
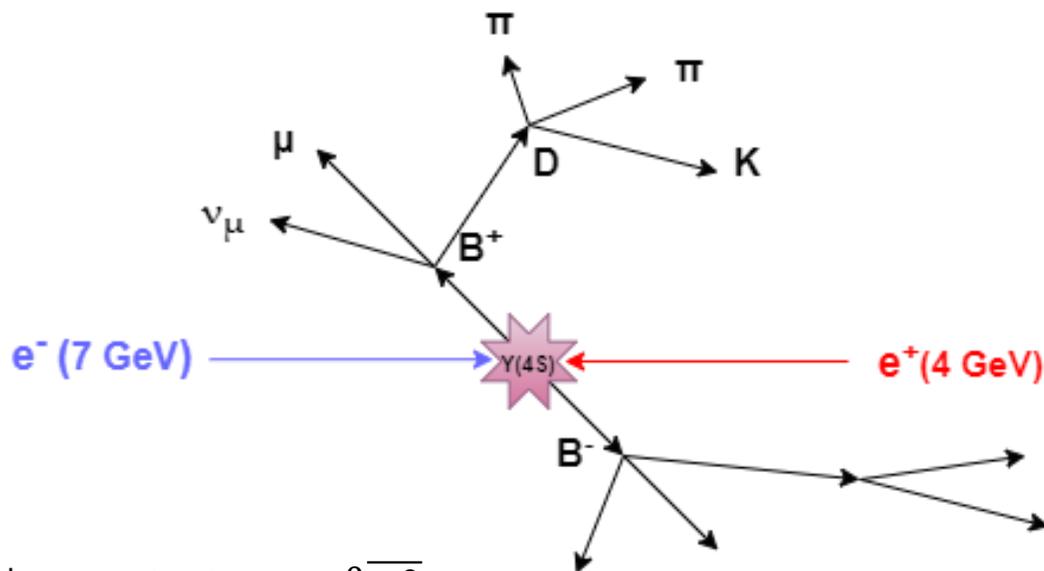
Contexte scientifique

- Nouvelle physique au-delà du modèle standard :
 - direct
 - indirect

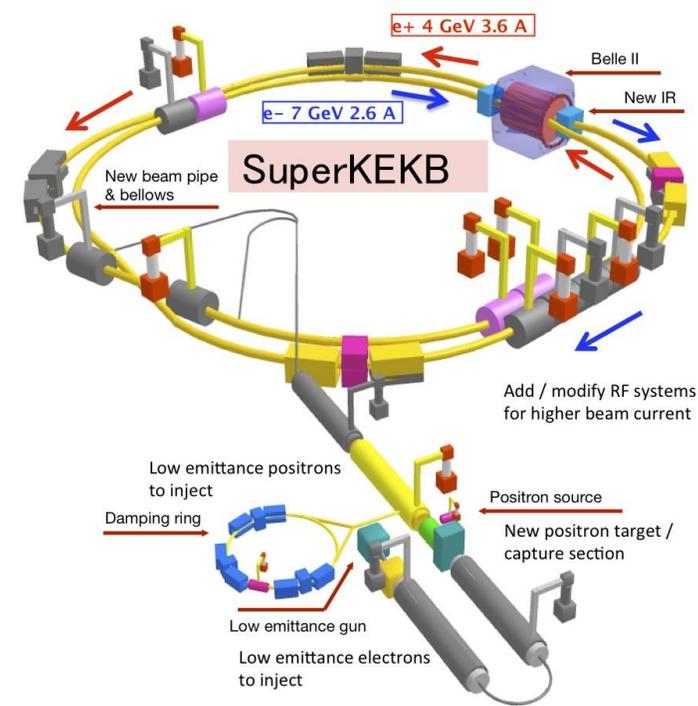
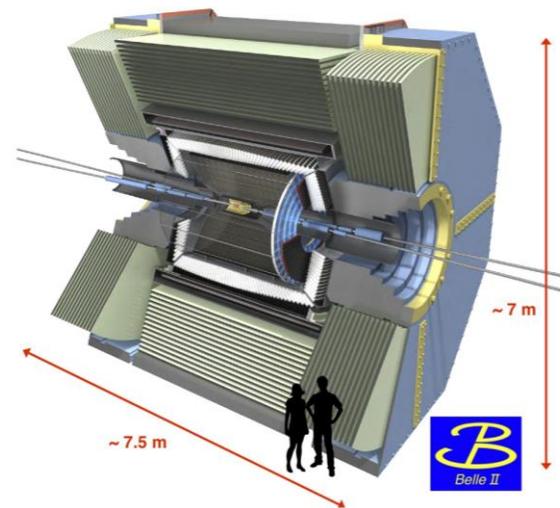


Belle II

- Expérience Belle II : désintégration du méson B



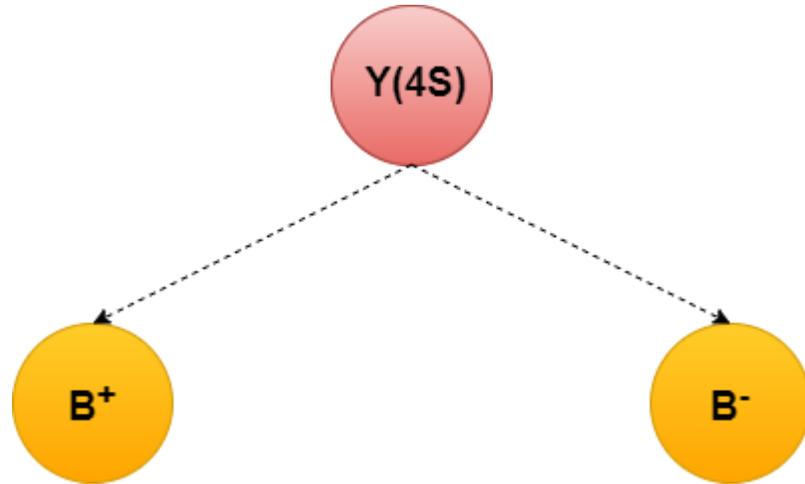
- SuperKEKB, Tsukuba (Japon) :
plus grande luminosité au monde en 2020 : $2,22 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$



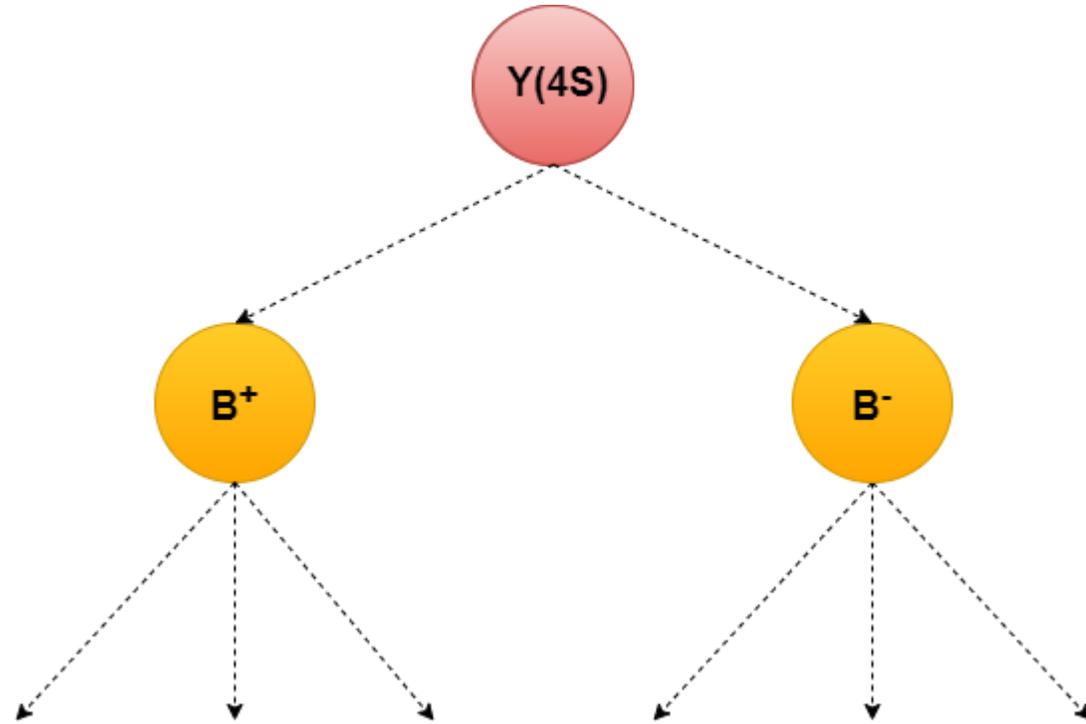
Cadre du projet



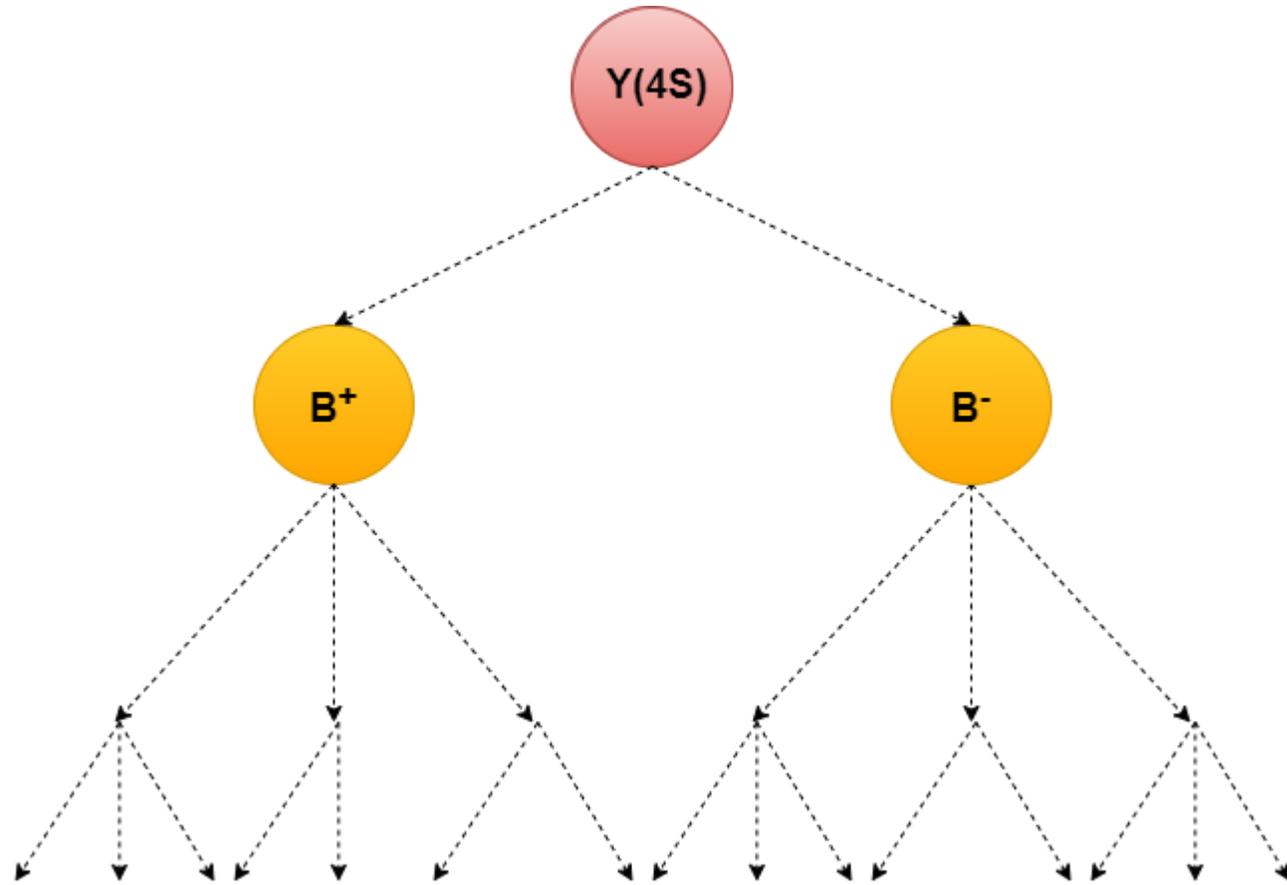
Cadre du projet



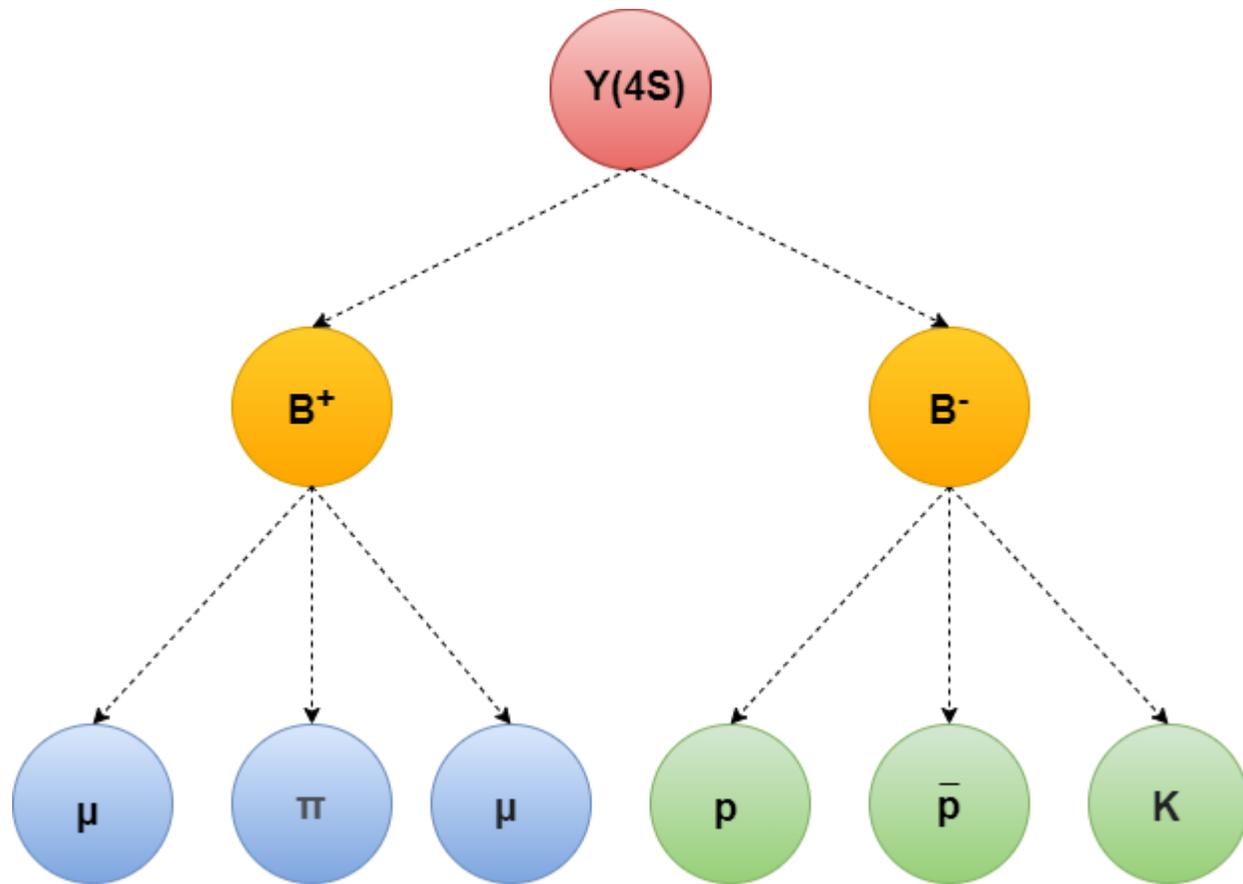
Cadre du projet



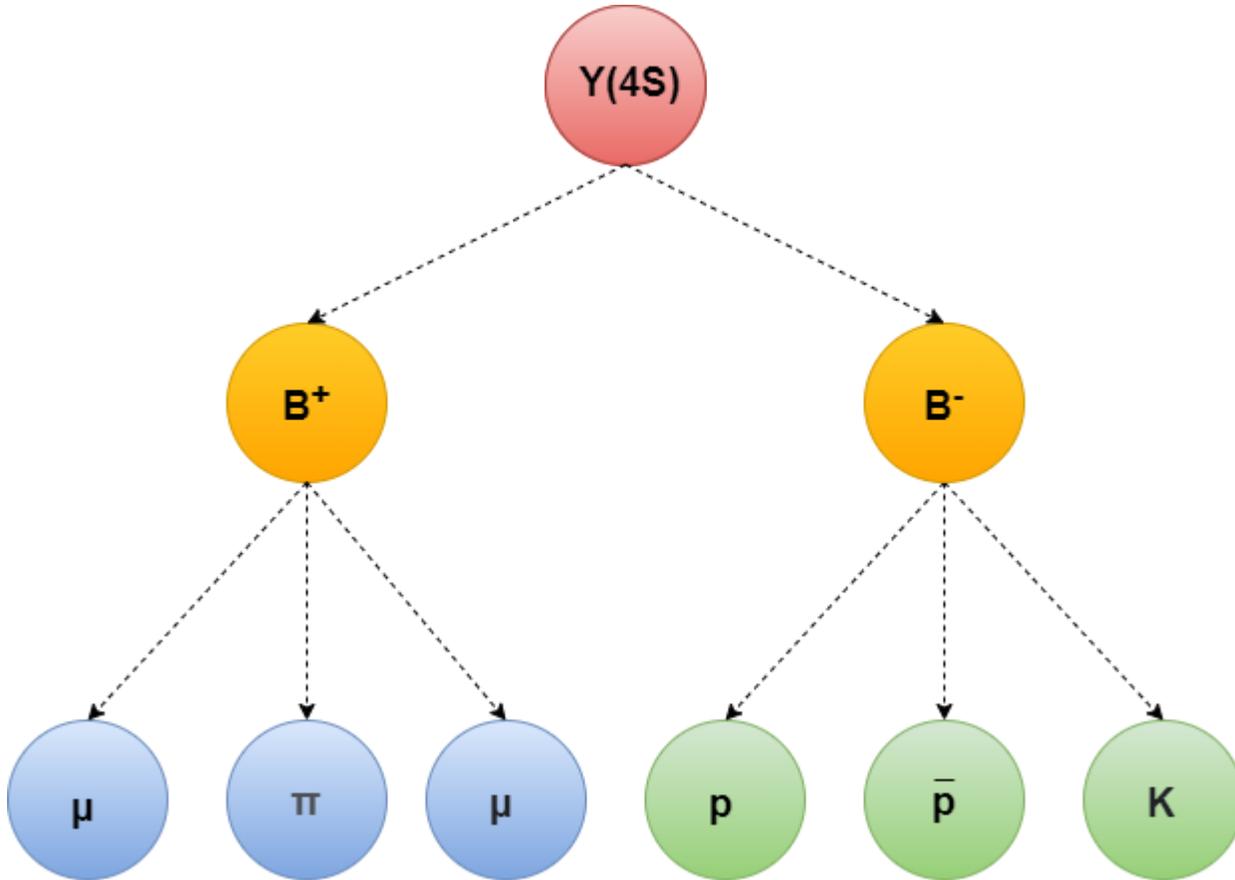
Cadre du projet



Cadre du projet



Cadre du projet



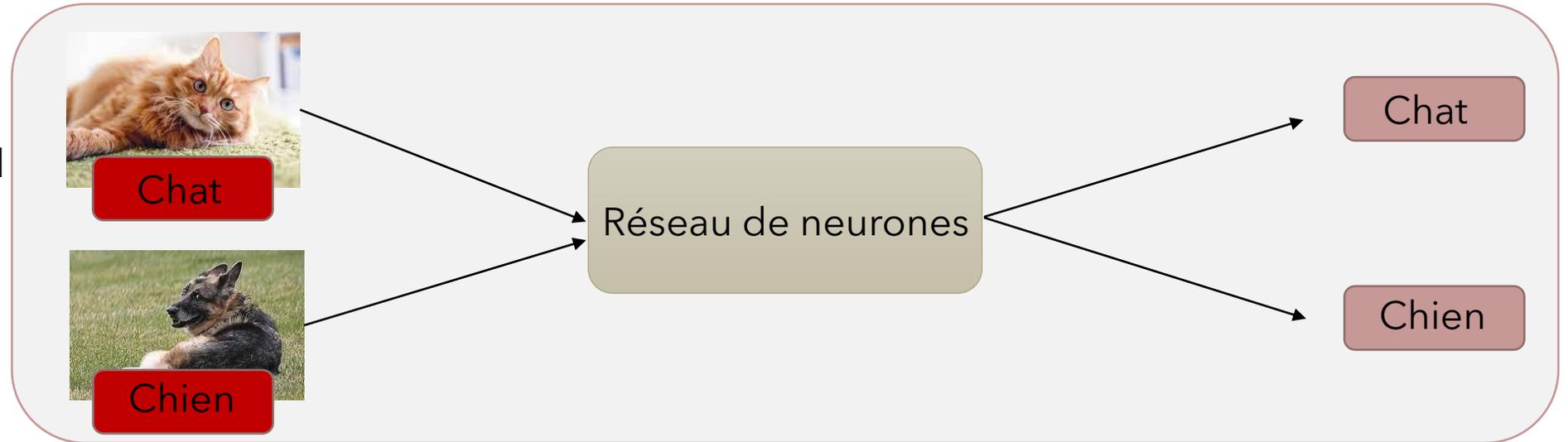
- **Objectif** : attribuer à chaque particule de l'état final le méson B duquel elle est issue
- **Cadre** : entre quatre et dix particules de l'état final, uniquement π , K , μ , e , p
- Topologie de type « 3o3 »
- Réseau de neurones à graphes



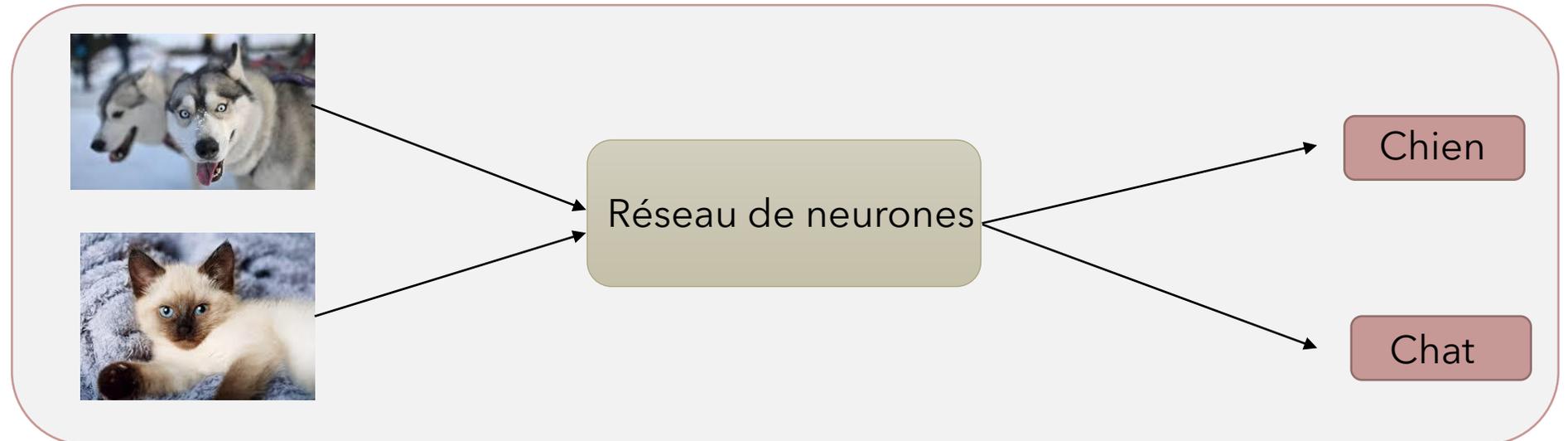
Réseau de neurones

Réseau de neurones

Entraînement :
le réseau apprend
à partir des
données qu'on lui
présente



Validation :
Le réseau
attribue lui
mêmes les
étiquettes

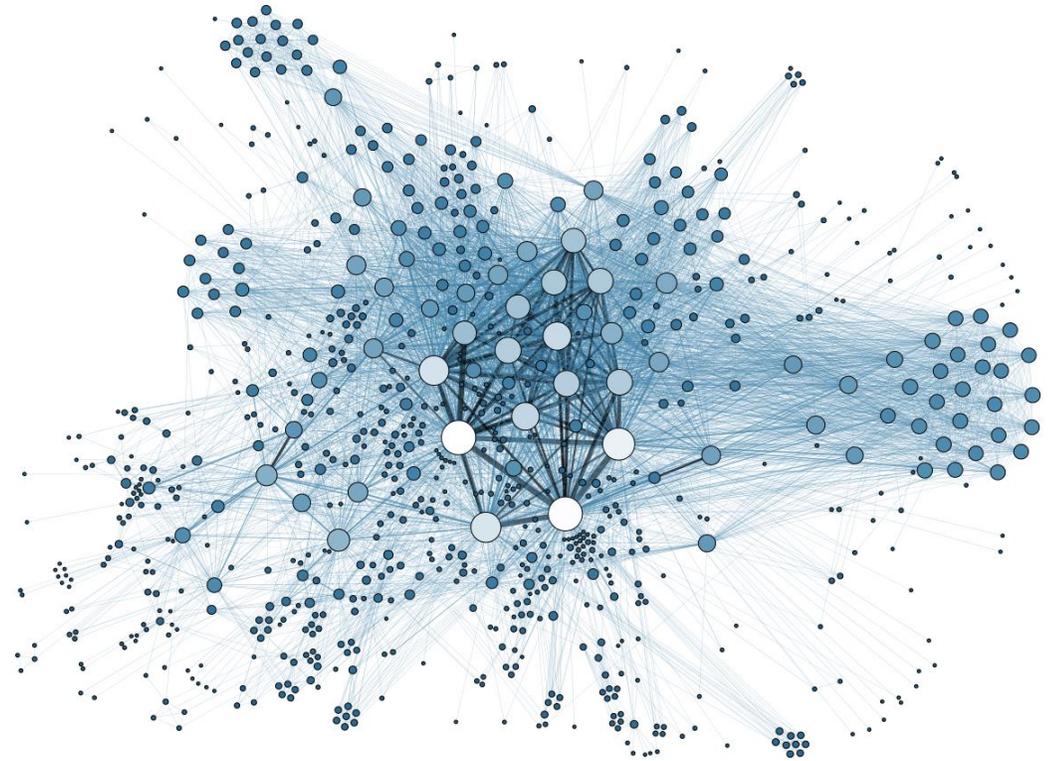
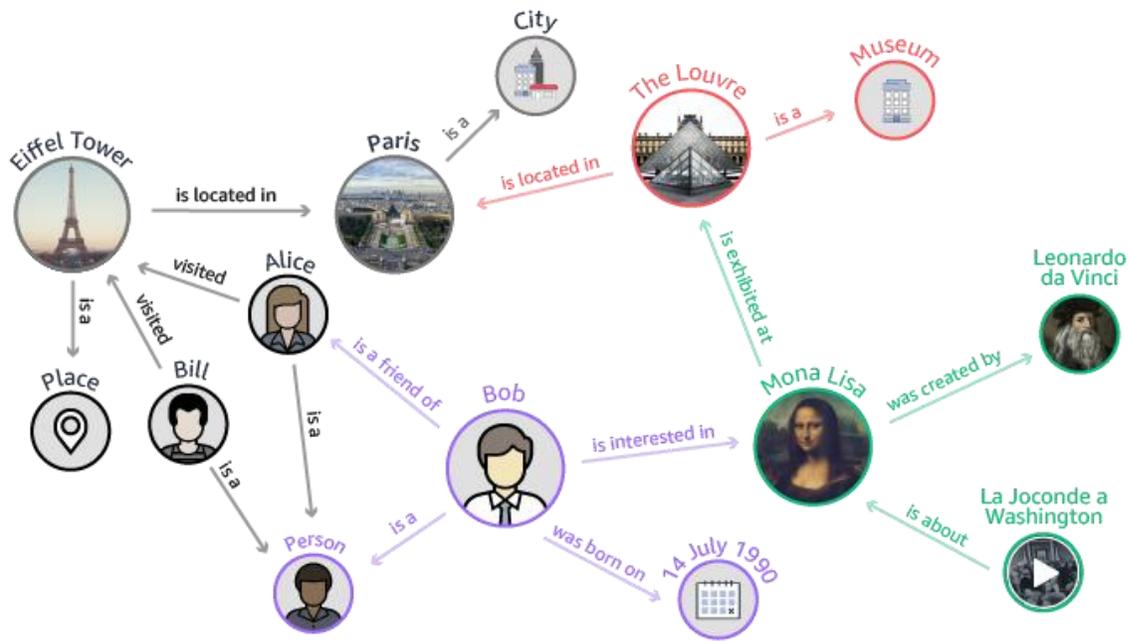




*Réseau de
neurons à
graphes*

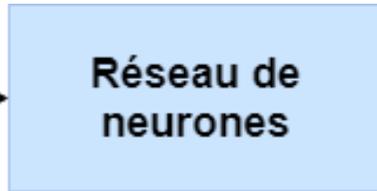
Graphes

- Graphe : ensemble de nœuds reliés par des liens



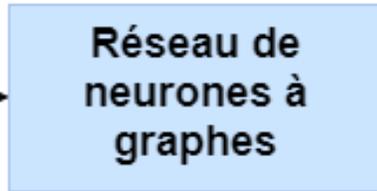
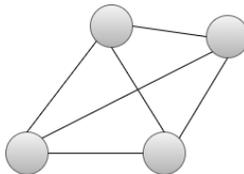
Réseau de neurones à graphes

Images



Prédictions

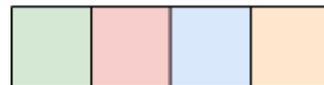
Matrices
et graphes



Prédictions

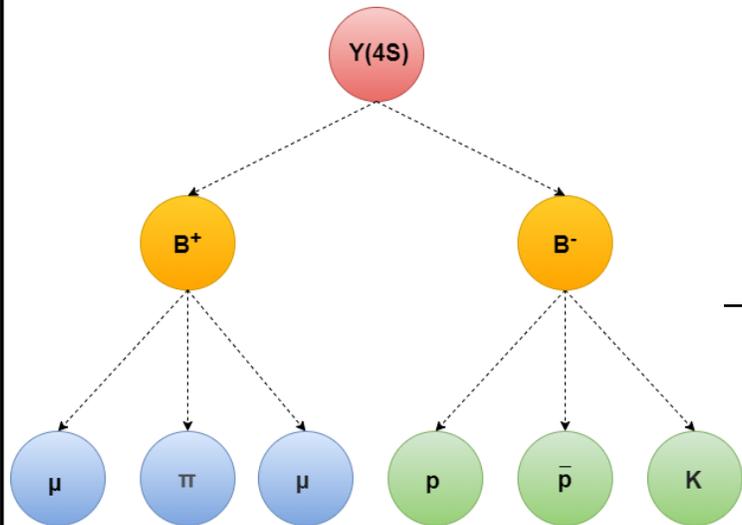


Attributs

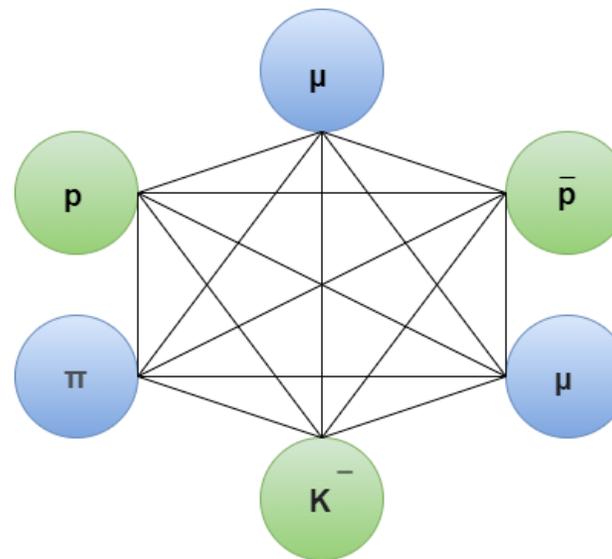


Réseau de neurones à graphes

Arbre de désintégration de $\Upsilon(4S)$



Grphe entièrement connecté entre les particules de l'état final

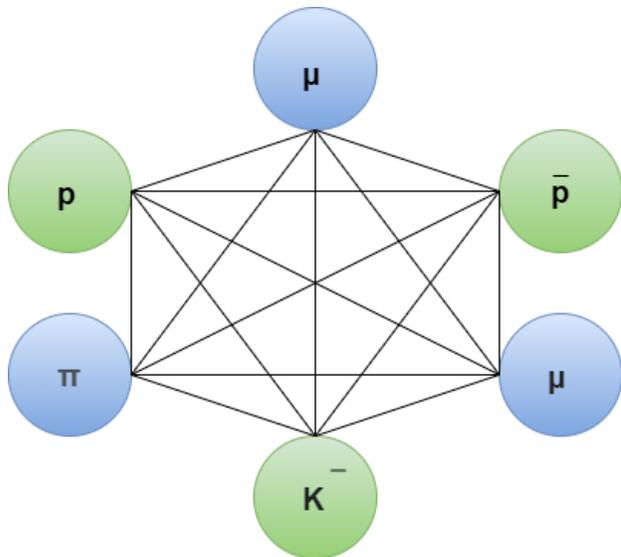


Matrice des relations

	μ	π	μ	p	\bar{p}	K
μ	0	1	1	2	2	2
π	1	0	1	2	2	2
μ	1	1	0	2	2	2
p	2	2	2	0	1	1
\bar{p}	2	2	2	1	0	1
K	2	2	2	1	1	0

Réseau de neurones à graphes

Graphe entièrement connecté entre les particules de l'état final



Matrice des quadri-impulsions

	P_x	P_y	P_z	E
μ	a_0	a_1	a_2	a_3
π	b_0	b_1	b_2	b_3
μ	c_0	c_1	c_2	c_3
p	d_0	d_1	d_2	d_3
\bar{p}	e_0	e_1	e_2	e_3
K	f_0	f_1	f_2	f_3



Attributs

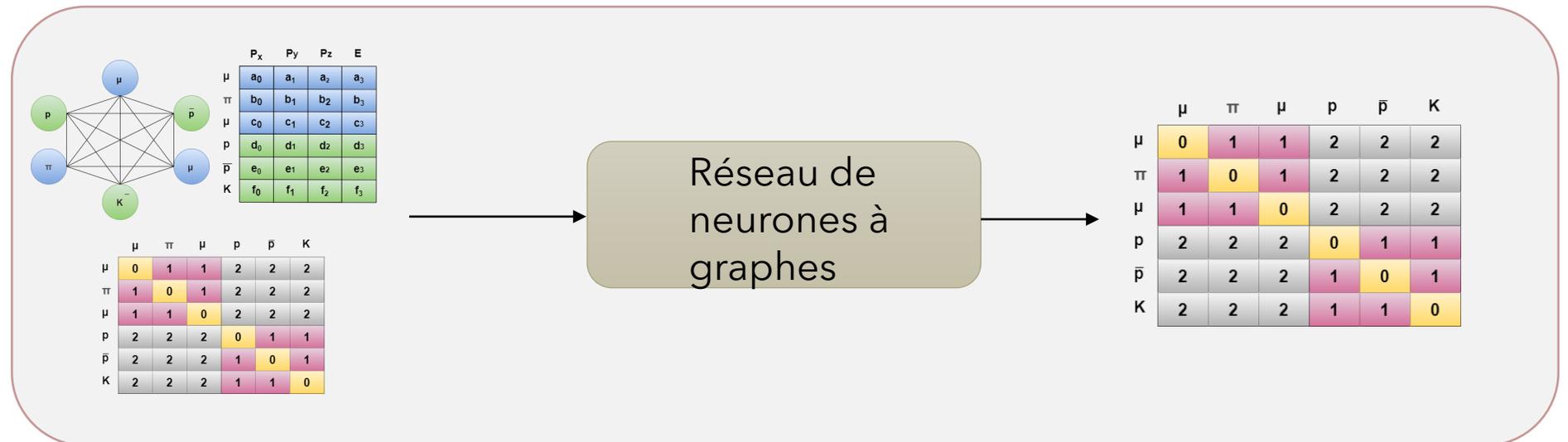
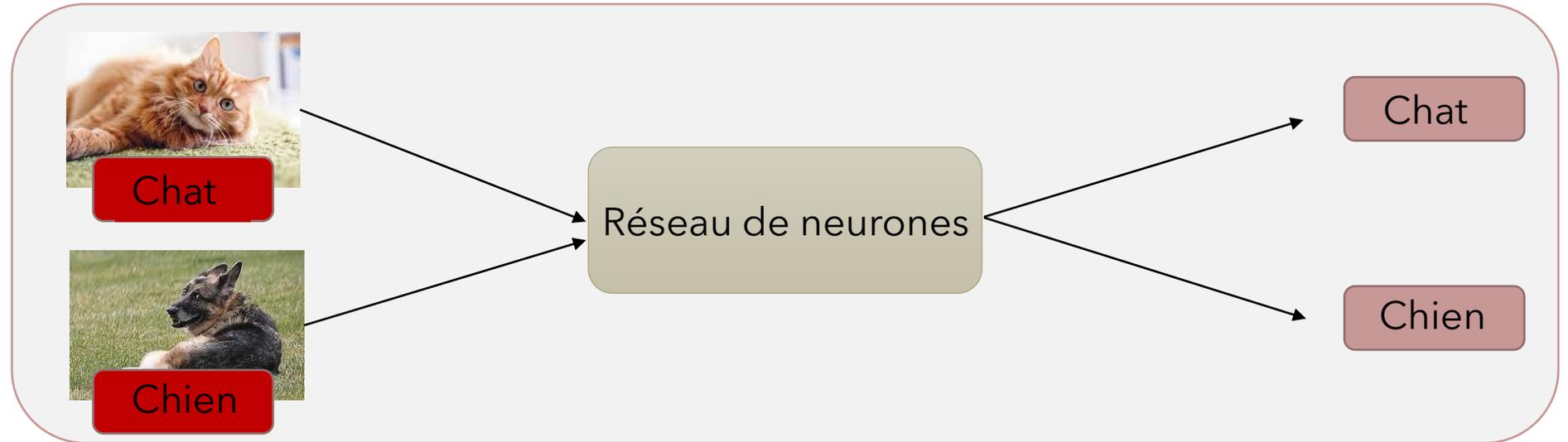


Matrice des relations

	μ	π	μ	p	\bar{p}	K
μ	0	1	1	2	2	2
π	1	0	1	2	2	2
μ	1	1	0	2	2	2
p	2	2	2	0	1	1
\bar{p}	2	2	2	1	0	1
K	2	2	2	1	1	0

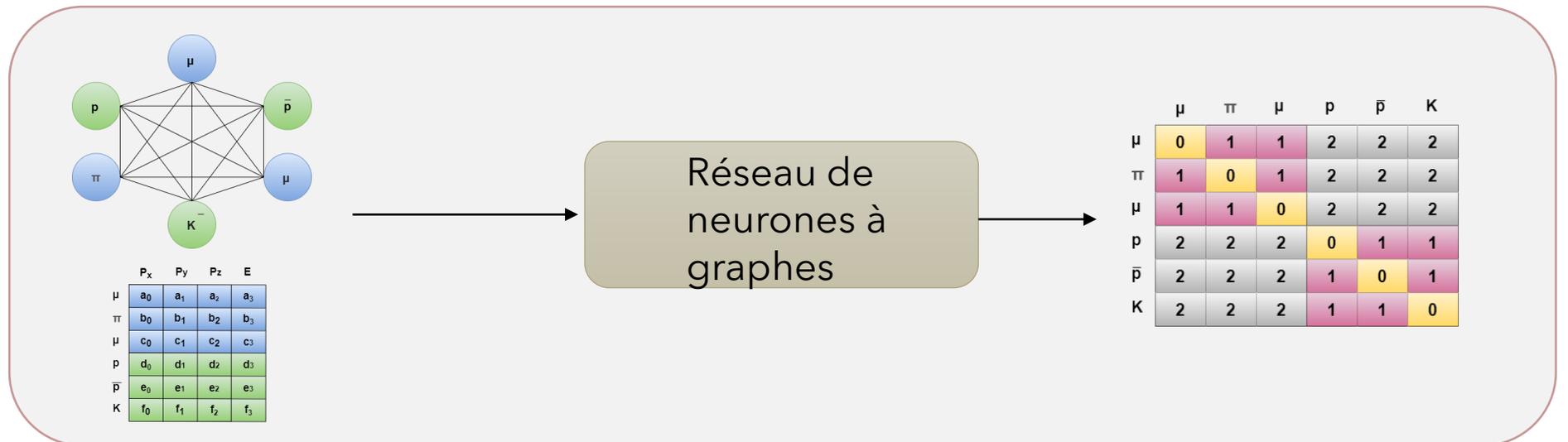
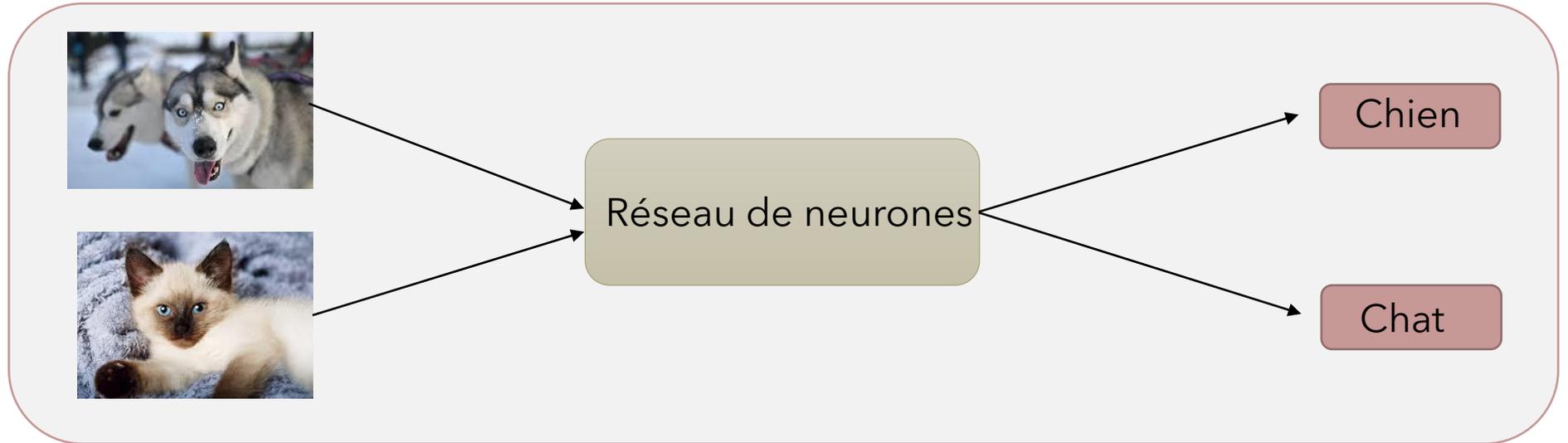
Réseau de neurones à graphes

Entraînement



Réseau de neurones à graphes

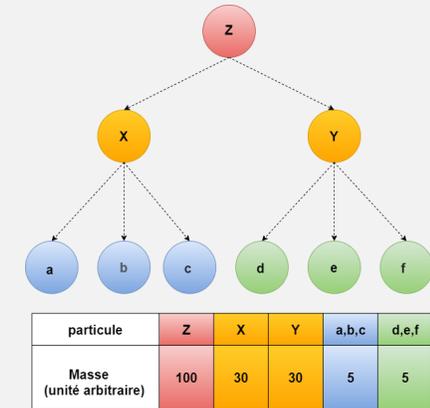
Validation



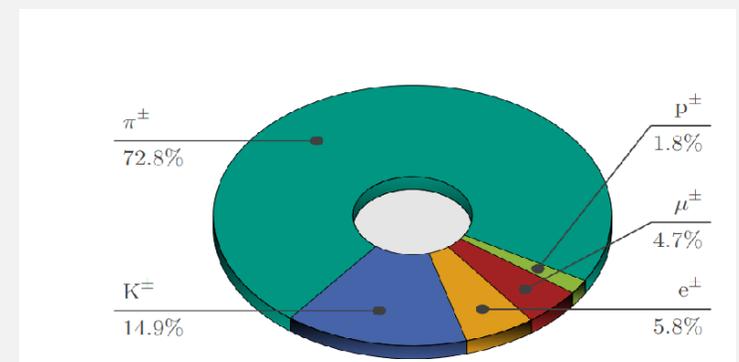
Données utilisées

➤ Bibliothèque phasespace

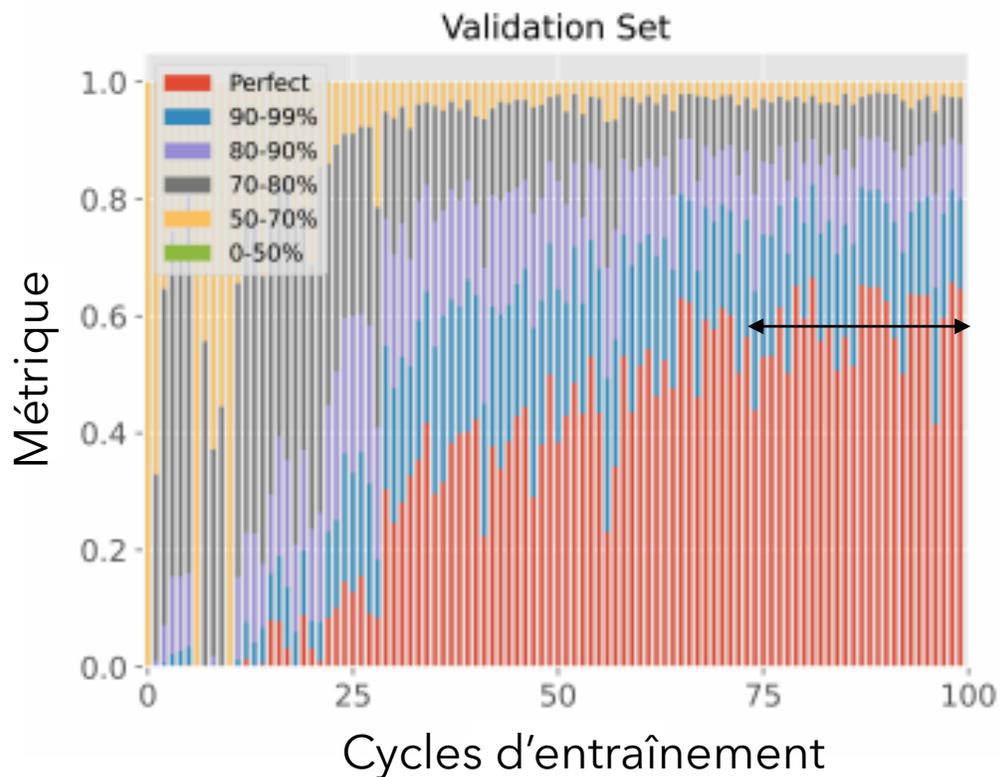
- Masses génériques, topologies « 3o3 », « 4o4 » et « 5o5 » :
 - Optimisation



- Masses réelles de la désintégration de $\Upsilon(4S)$:
 - $\Upsilon(4S)$: 10,579 GeV
 - mésons B : 5,2798 GeV
 - FSP : proportions réellement observées
 - topologies : entre quatre et dix
 - augmentation complexité



Résultats & discussion



➤ Métrique, matrices des relations prédites :

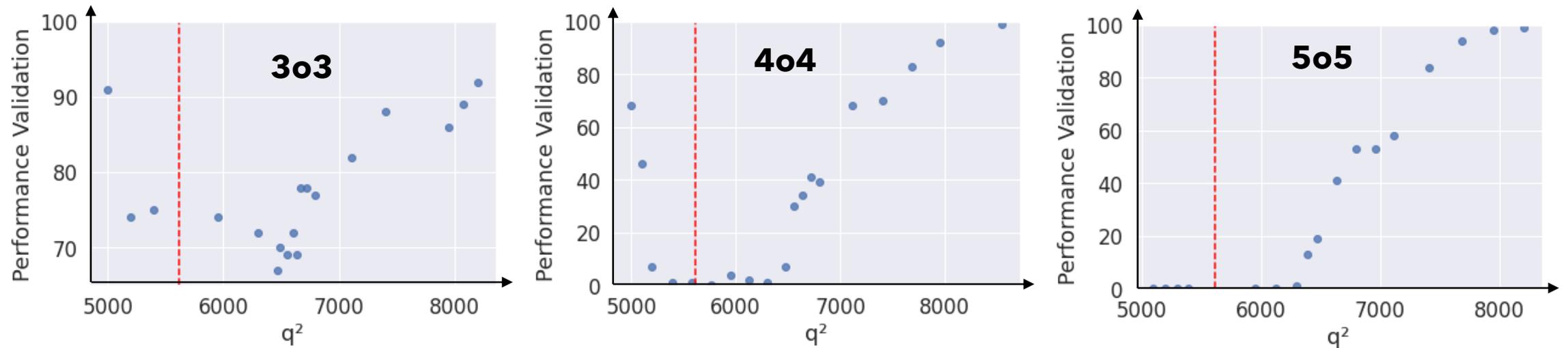
- parfaitement
- de 90 à 99%
- de 80 à 90%
- de 70 à 80%
- de 50 à 70%
- de 0 à 50%

←→ **Performance validation** :
proportion moyenne des matrices des relations parfaitement prédites sur les 25 dernières époques

Cycle d'entraînement :
nombre d'itérations sur les données durant l'entraînement

Résultats & discussion

➤ Masses génériques

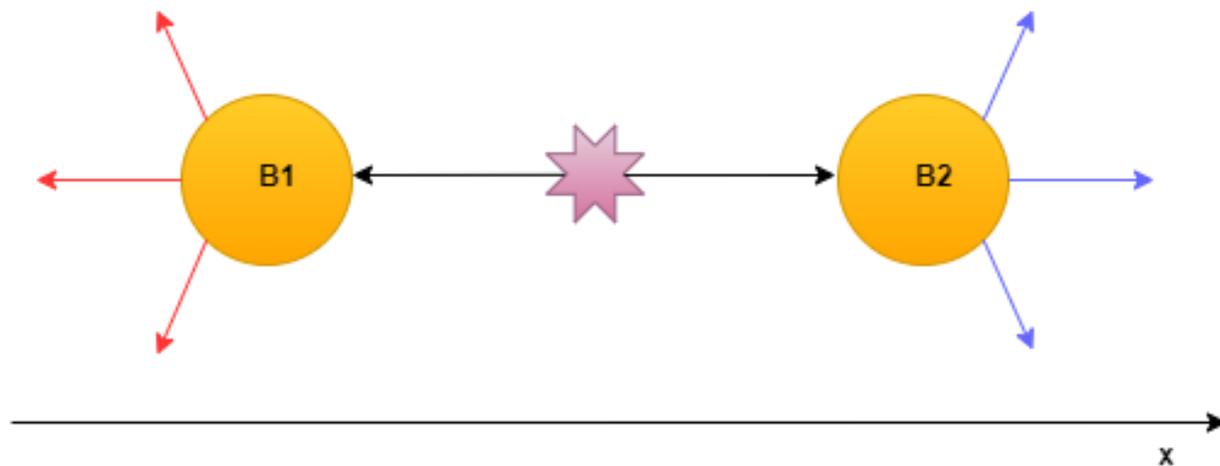
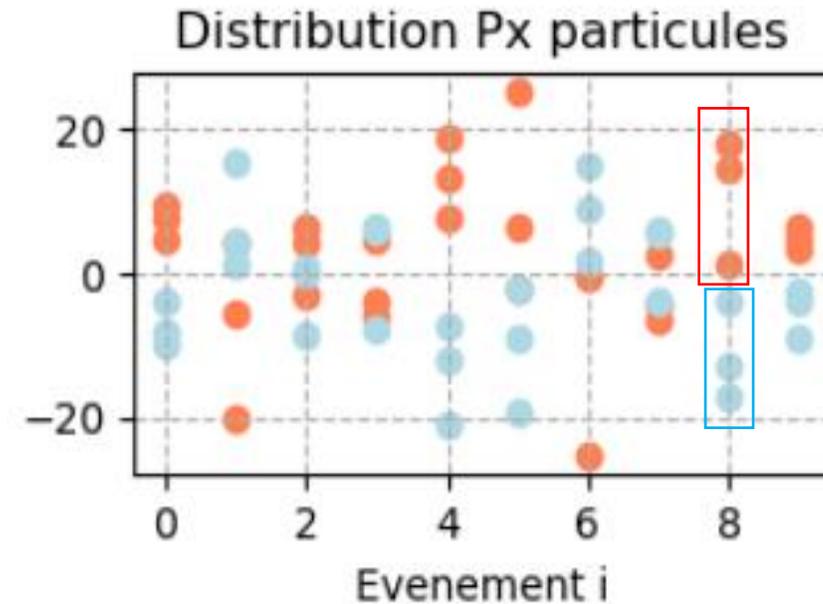


- Énergie de transfert : $q^2 = \text{masse}(\text{particule racine})^2 - \text{masse}(\text{droite})^2 - \text{masse}(\text{gauche})^2$
- Grande énergie de transfert → apprentissage grâce aux quadri-impulsions des particules de l'état final

Résultats & discussion

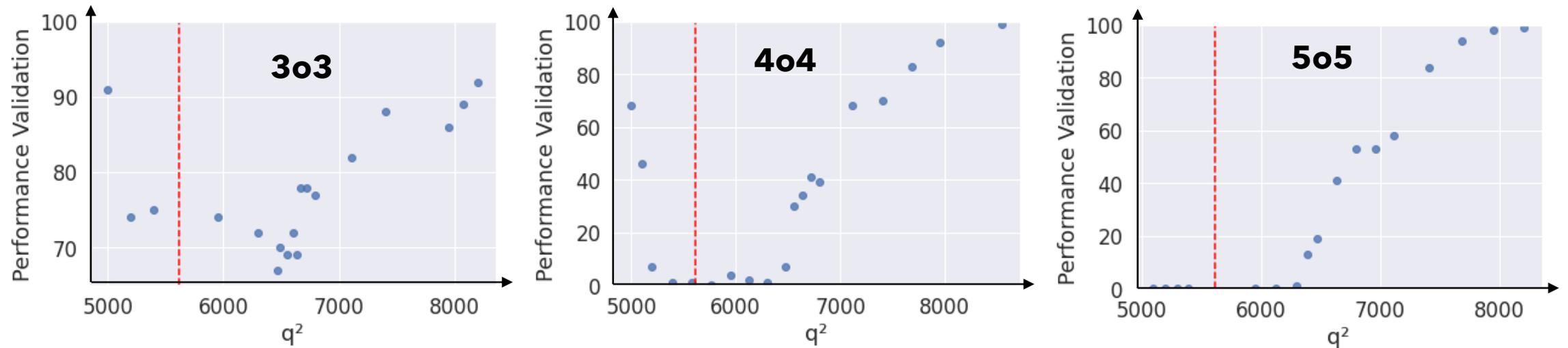
Distribution des quadri-impulsion particules de l'état final

- $q^2 = 8200 \text{ MeV}^2$
- échantillon de 10 évènements
- topologie 3o3



Résultats & discussion

➤ Masses génériques

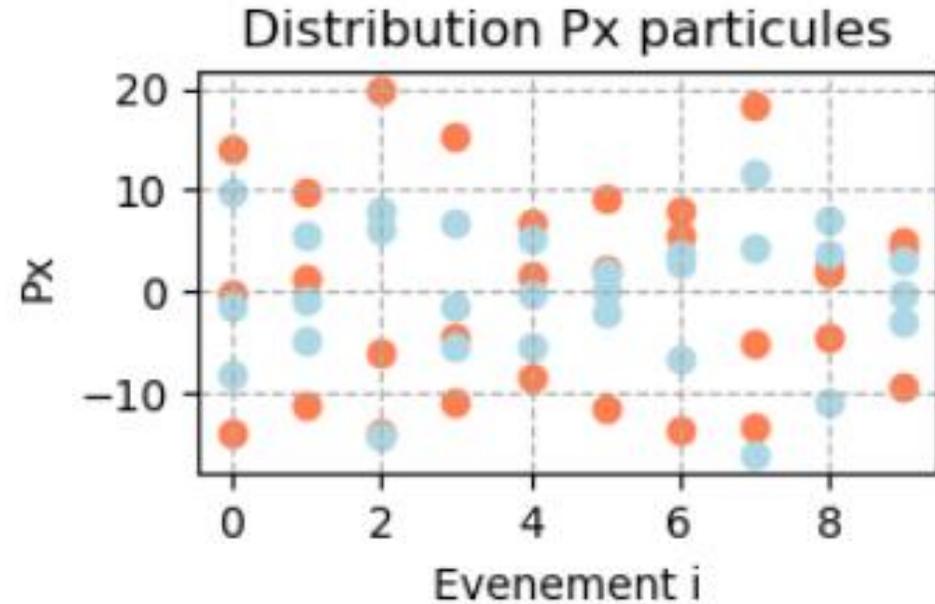
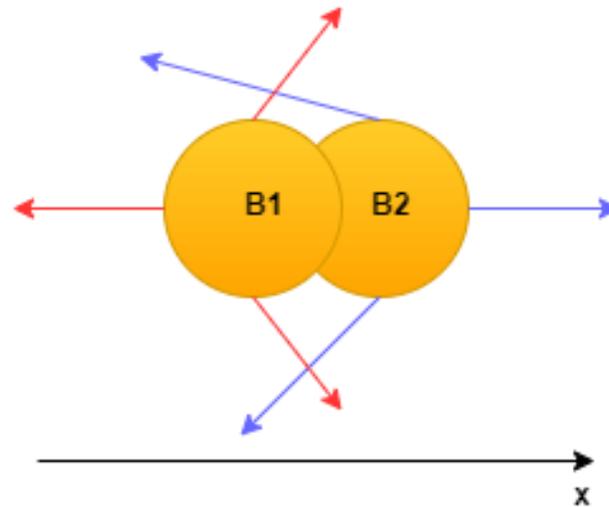


- Énergie de transfert : $q^2 = \text{masse}(\text{particule racine})^2 - \text{masse}(\text{droite})^2 - \text{masse}(\text{gauche})^2$
- Grande énergie de transfert → apprentissage grâce aux quadri-impulsions des particules de l'état final
- Petite énergie de transfert → apprentissage grâce aux masses (combinaisons)

Résultats & discussion

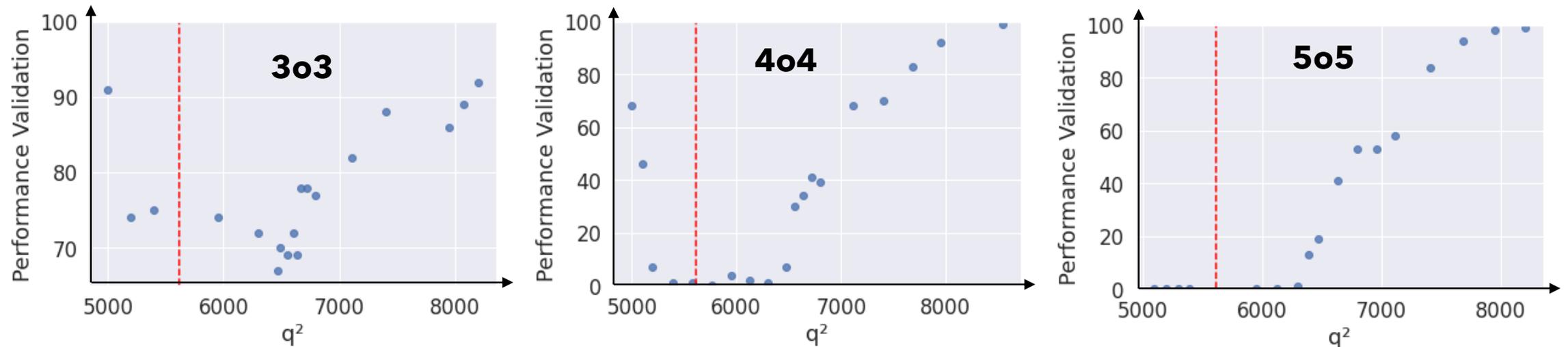
Distribution des quadri-impulsion particules de l'état final

- $q^2 = 5000 \text{ MeV}^2$
- échantillon de 10 évènements
- topologie 3o3



Résultats & discussion

➤ Masses génériques



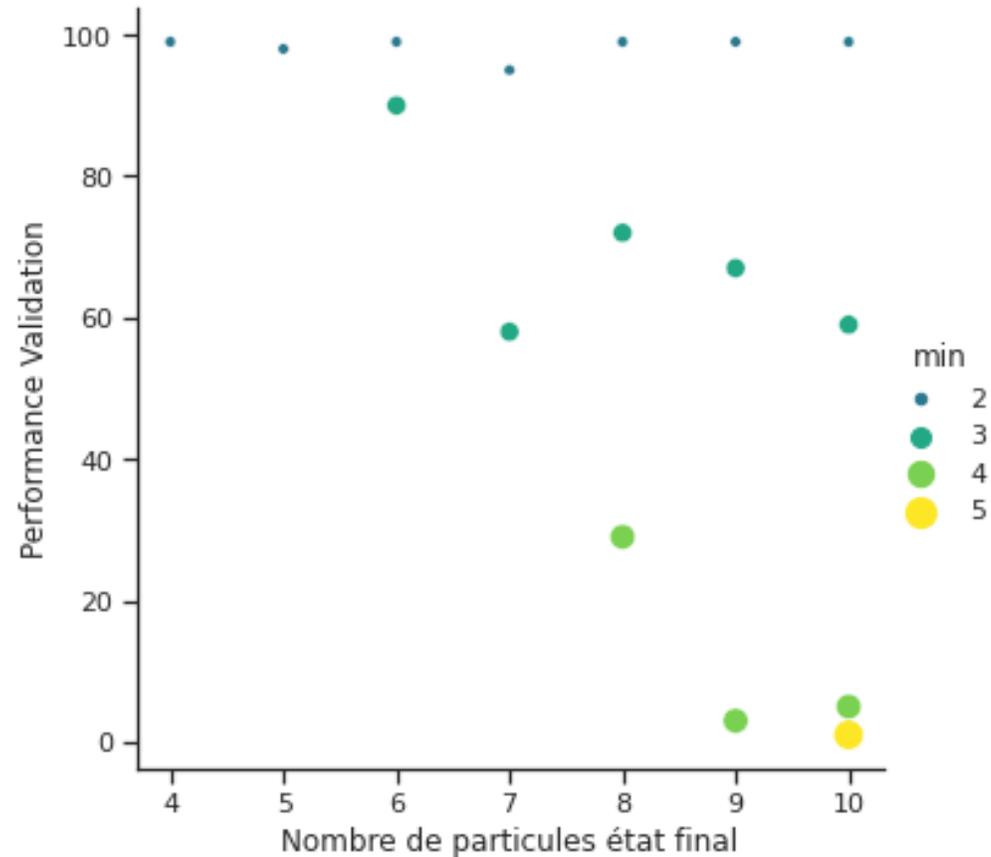
- Énergie de transfert : $q^2 = \text{masse}(\text{particule racine})^2 - \text{masse}(\text{droite})^2 - \text{masse}(\text{gauche})^2$
- Grande énergie de transfert → apprentissage grâce aux quadri-impulsions des particules de l'état final
- Petite énergie de transfert → apprentissage grâce aux masses (combinaisons)

Résultats & discussion

- Masses réelles de la désintégration de $\Upsilon(4S)$: une seule topologie

Exemple 4o2 :

Nombre de particules de l'état final : 6
min : 2

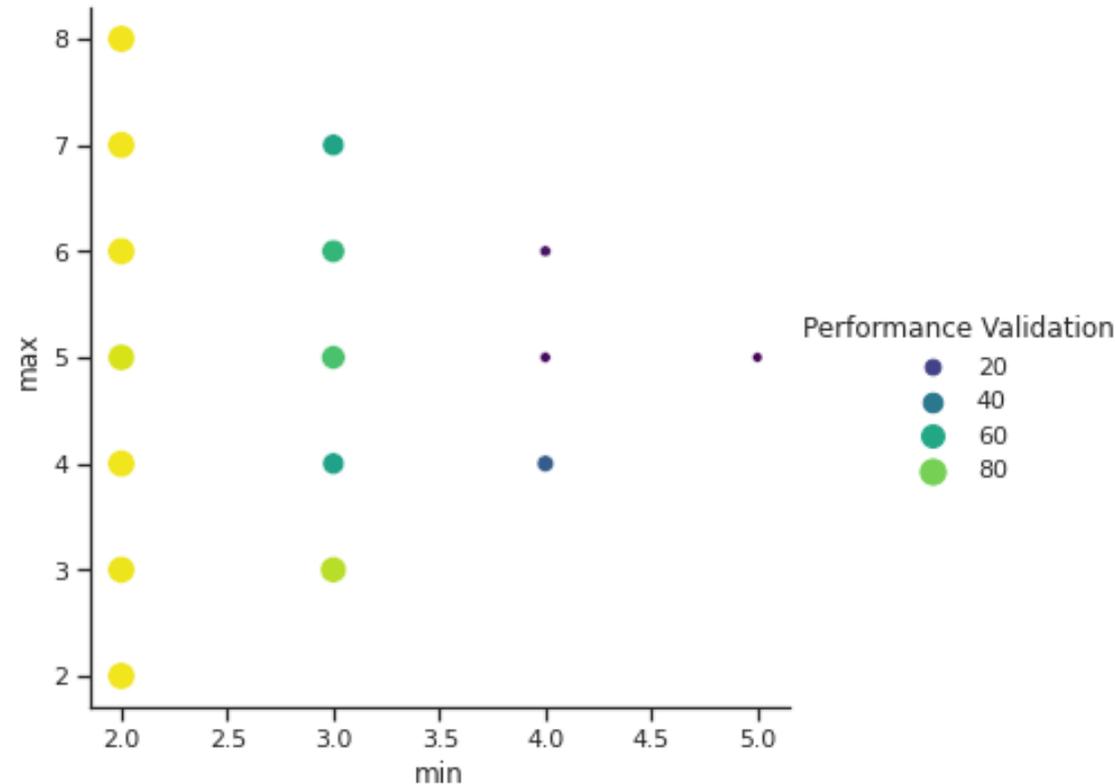


Résultats & discussion

- Masses réelles de la désintégration de $\Upsilon(4S)$: une seule topologie

Exemple 4o2 :

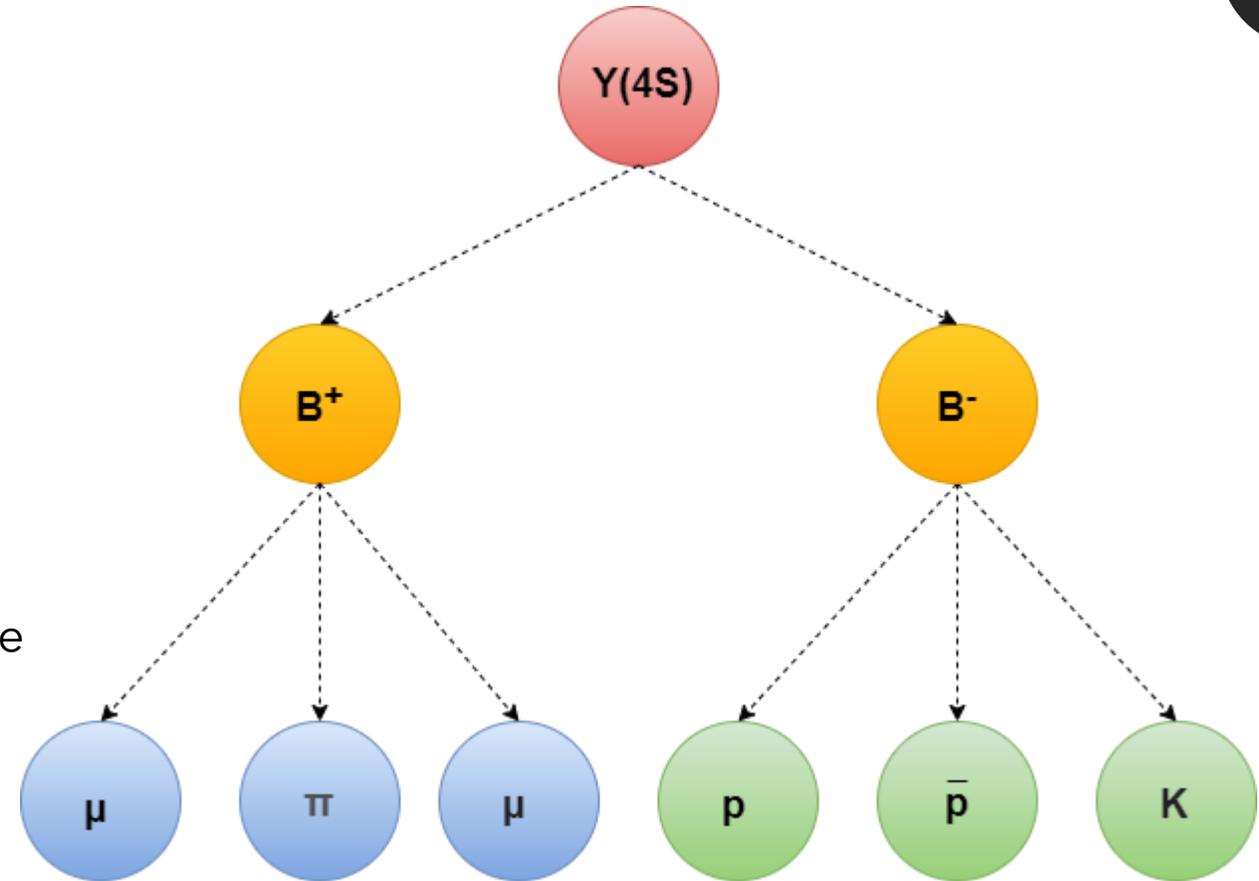
Nombre de particules de l'état final : 6
min : 2



- Minimum a la plus grande influence
- 2 particules issues d'un même méson B : désintégration triviale

Conclusion

- Optimisation
- Premiers tests réalisés :
 - Masses génériques
 - Masses réelles avec une seule topologie



Perspectives

- Entraînements avec les données Monte Carlo de Belle II

- plusieurs topologies

Nombre FSP	Topologies	Performance Training	Performance Validation
4, 5 et 6	2o2, 3o2, 3o3, 4o2	89.2%	86.7%
4, 5, 6 et 7	2o2, 3o2, 3o3, 4o2, 3o4, 5o2	91.1%	73.1%
4, 5, 6, 7 et 8	2o2, 3o2, 3o3, 4o2, 3o4, 5o2, 3o5, 4o4, 6o2	89.4%	77.0%
4, 5, 6, 7, 8 et 9	2o2, 3o2, 3o3, 4o2, 3o4, 5o2, 3o5, 4o4, 6o2, 3o6, 4o5, 7o2	75.7%	70.4%
4, 5, 6, 7, 8, 9 et 10	2o2, 3o2, 3o3, 4o2, 3o4, 5o2, 3o5, 4o4, 6o2, 3o6, 4o5, 7o2, 3o7, 6o4, 5o5, 8o2	72.1%	65.9%

- Améliorations possibles :

- Prise en compte des particules « manquantes » comme les neutrinos

- Prise en compte du bruit

Merci pour votre attention

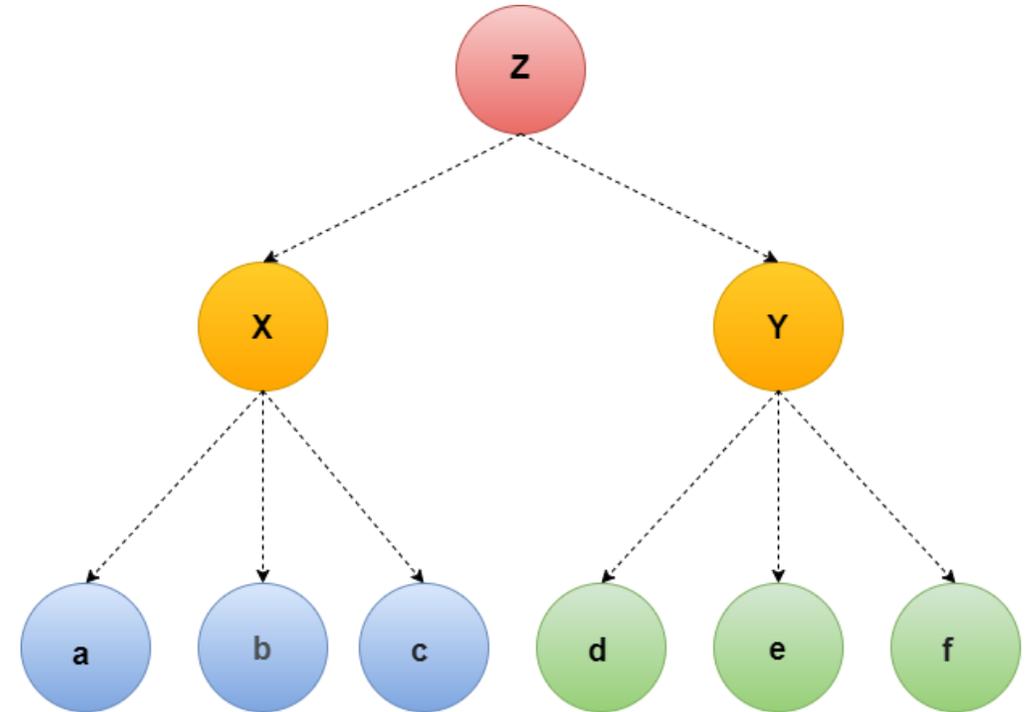
Questions?

Bibliographie

- [1] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [2] W. Altmannshofer et al. The Belle II Physics Book. *PTEP*, 2019(12) :123C01, 2019. [Erratum : *PTEP* 2020, 029201 (2020)].
- [3] A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J. C*, 74 :3026, 2014.
- [4] Yuki Yoshi Ohnishi et al. Accelerator design at SuperKEKB. *PTEP*, 2013 :03A011, 2013.
- [5] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. SuperKEKB Collider. *Nucl. Instrum. Meth. A*, 907 :188–199, 2018.
- [6] Jayashree Kalpathy-Cramer Michael F. Chiang Rene Y. Choi, Aaron S. Coyner and J. Peter Campbell. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol.*, 2020 Feb 27.
- [7] I.H Sarker. Machine Learning : Algorithms, Real-World Applications and Research Directions. *SN Computer Science volume*, 22 March 2021.
- [8] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph Neural Networks in Particle Physics. 7 2020.
- [9] Peter W. Battaglia et al. Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261, 2018.
- [10] PyTorch, <https://pytorch.org>.
- [11] Documentation du module Geometric de Pytorch, <https://pytorch-geometric.readthedocs.io>.
- [12] Documentation de phasespace, <https://pypi.org/project/phasespace/>.
- [13] Valerio Bertacchi et al. Track finding at Belle II. *Comput. Phys. Commun.*, 259 :107610, 2021.
- [14] Cross Entropy, <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>.
- [15] Adam, <https://pytorch.org/docs/stable/optim.html>.

Hyperparamètres

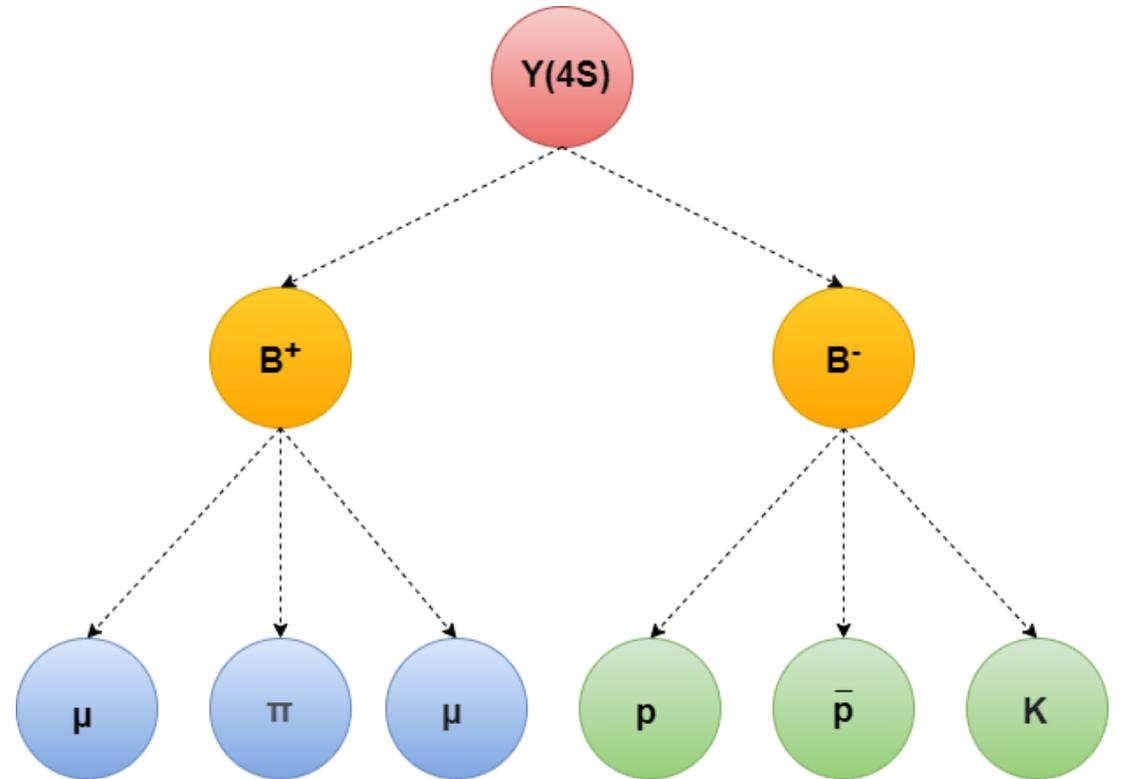
- Masses génériques:
 - nombre d'époques: 100
 - groupement de données: 20
 - taux d'apprentissage: 0,001
 - nombre d'évènements: 5000



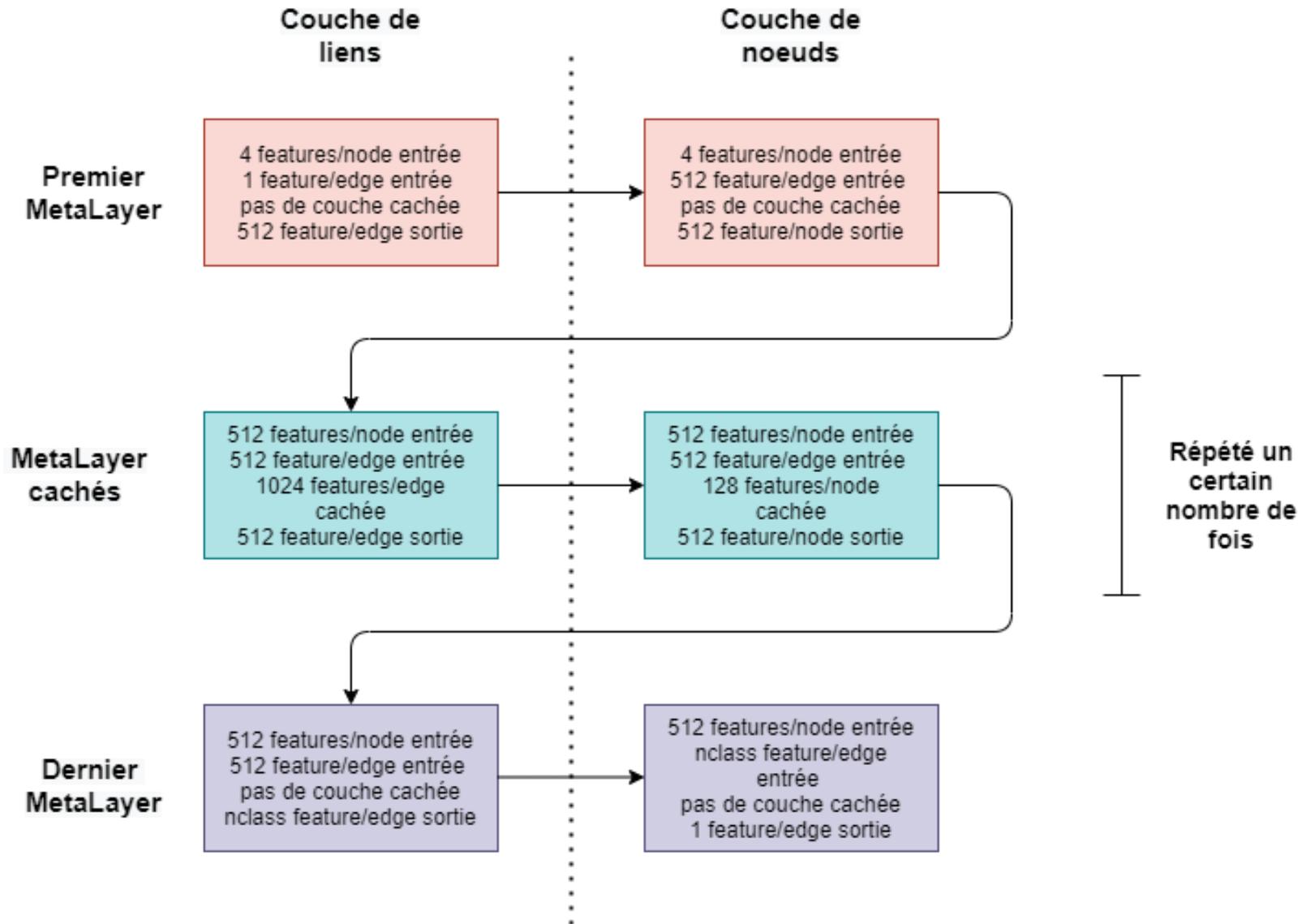
particule	Z	X	Y	a,b,c	d,e,f
Masse (unité arbitraire)	100	30	30	5	5

Hyperparamètres

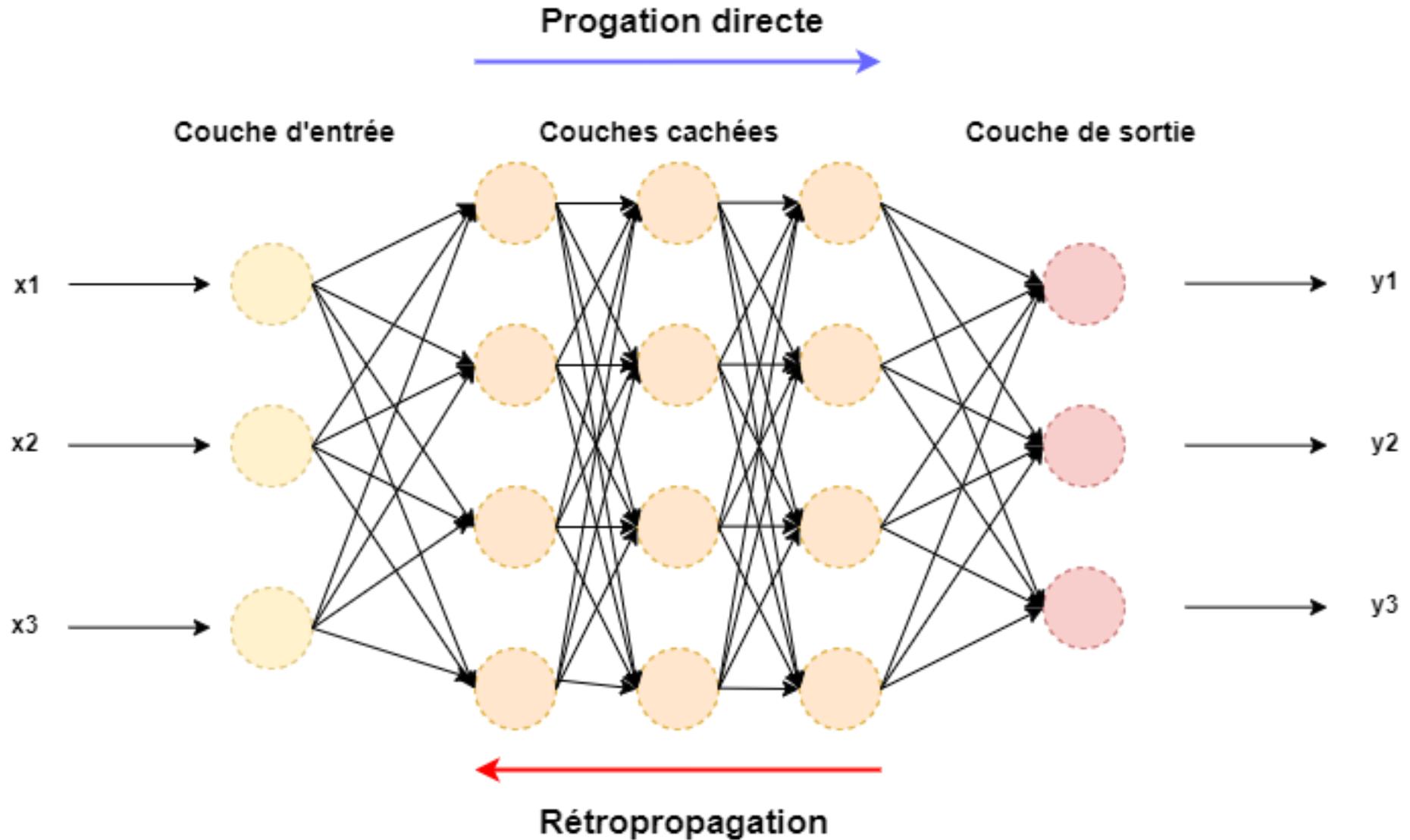
- Masses génériques:
 - nombre d'époques: 100
 - groupement de données: 60
 - taux d'apprentissage: 0,001
 - nombre d'évènements: 5000, puis 2500



Annexes

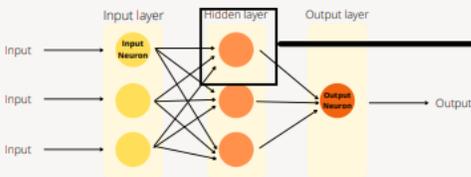


Annexes

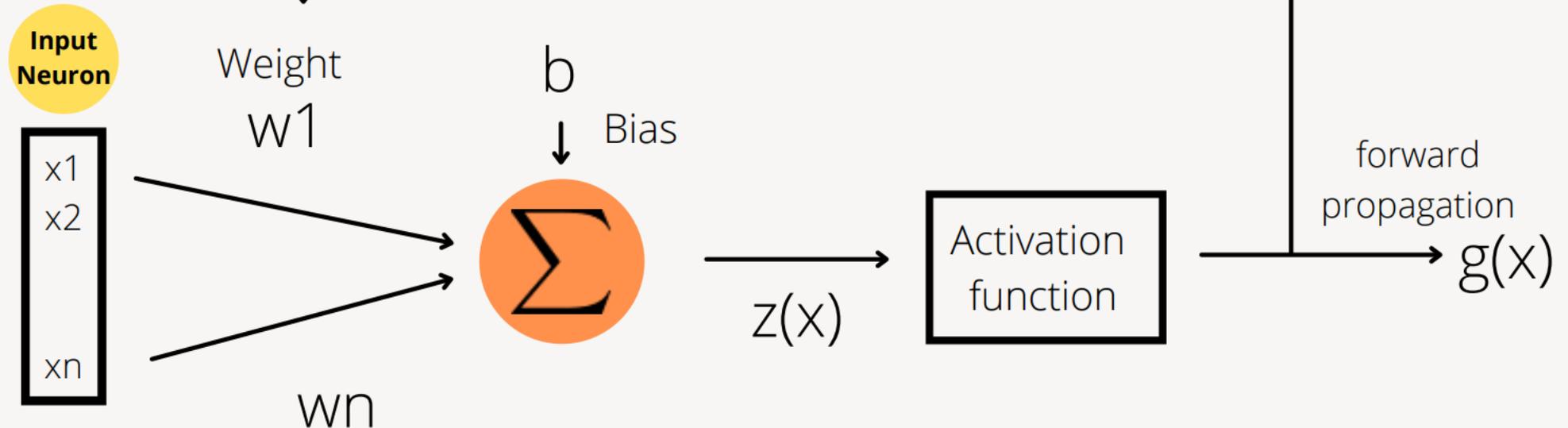


Réseau de neurones

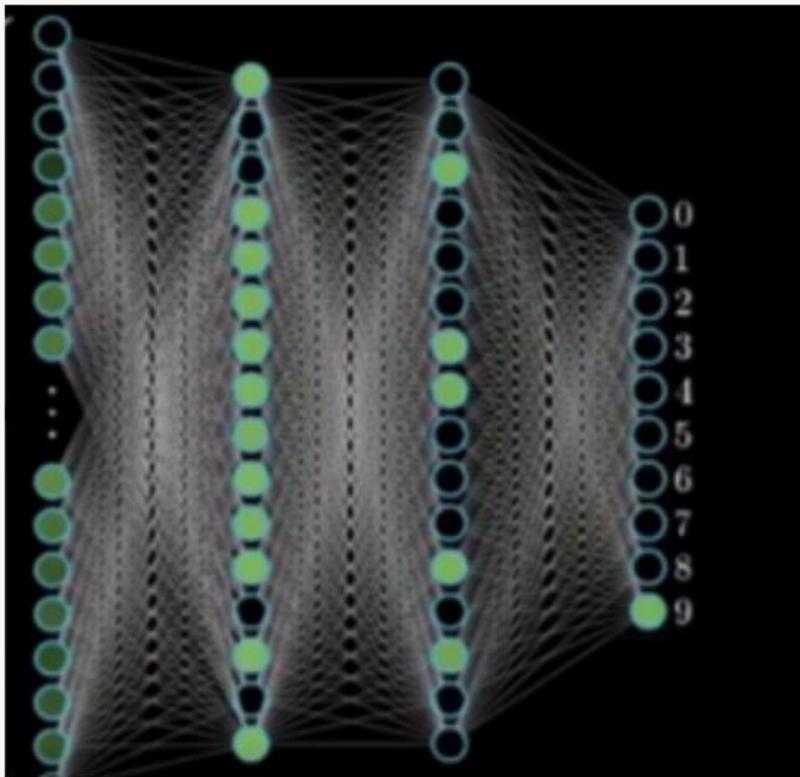
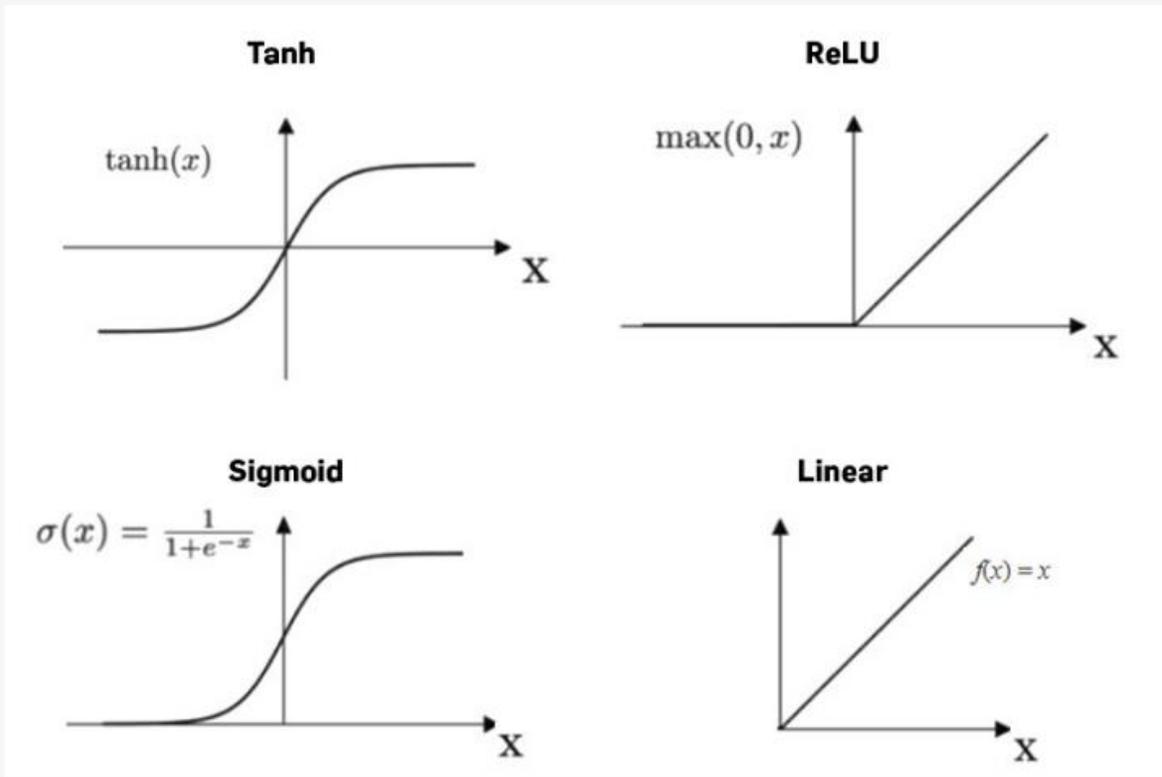
≡ Forward/back propagation



- Linear function: $z=Wx+b$
- Activation function: $a=g(x)$



Réseau de neurones



Neurone

$$\mathcal{N}_{\mathbf{w},b}(\mathbf{x}) = \sum_i w_i \cdot x_i + b$$

$$= \begin{bmatrix} w_0 & \cdots & w_m \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_m \end{bmatrix} + b$$

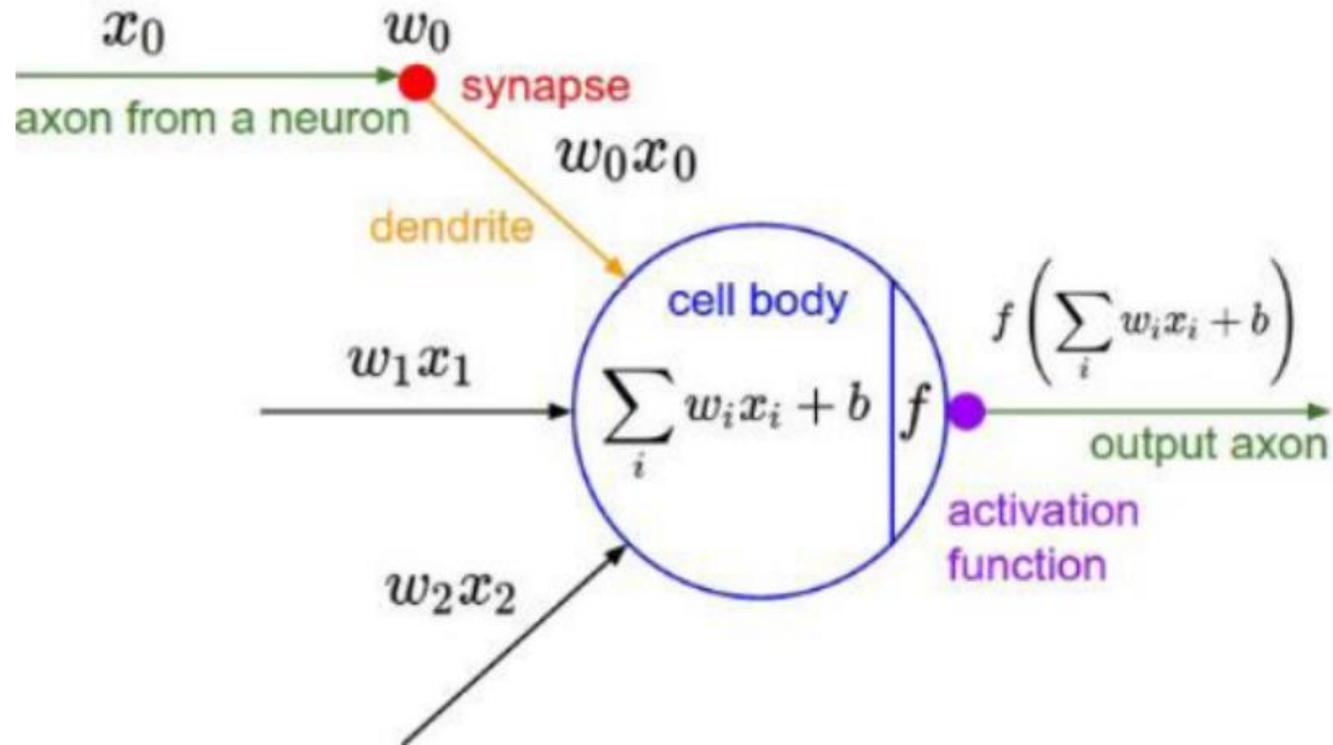
$$= \mathbf{w}^T \mathbf{x} + b$$

$$= z$$

Couche de neurones

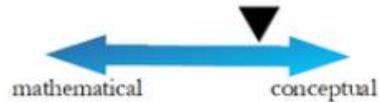
$$\begin{aligned}\mathcal{L}_{\mathbf{W},\mathbf{b}}(\mathbf{x}) &= \begin{bmatrix} \sum_i W_{i,0} \cdot x_i + b_0 \\ \vdots \\ \sum_i W_{i,n} \cdot x_i + b_n \end{bmatrix} \\ &= \begin{bmatrix} W_{0,0} & \cdots & W_{m,0} \\ \vdots & & \vdots \\ W_{0,n} & \cdots & W_{m,n} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ \vdots \\ x_m \end{bmatrix} + \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix} \\ &= \mathbf{W}^T \mathbf{x} + \mathbf{b} \\ &= \mathbf{z}\end{aligned}$$

Réseau de neurones

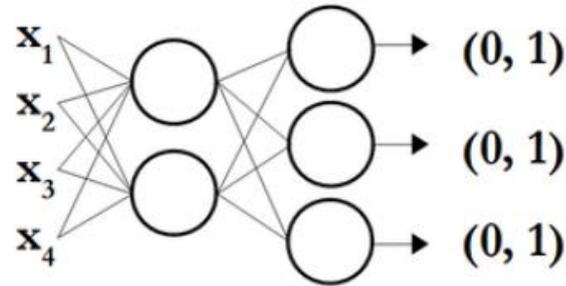


Réseau de neurones

How does the algorithm make a decision?



Feedforward



$$H_{in} = XW_1$$

$$H_{out} = \text{sigmoid}(H_{in})$$

$$O_{in} = H_{out}W_2$$

$$O_{out} = \text{sigmoid}(O_{in})$$

How do you determine the right parameters for the algorithm?



Cost function/Loss function:

$$\text{Mean Squared Error} = \frac{1}{2N} \sum_e \sum_n (O_{out\ e,n} - Y_{e,n})^2$$

Gradient Descent:

$$W_i := W_i - \alpha \cdot \frac{\partial \text{MSE}}{\partial W_i}(W_i)$$

Chain Rule:

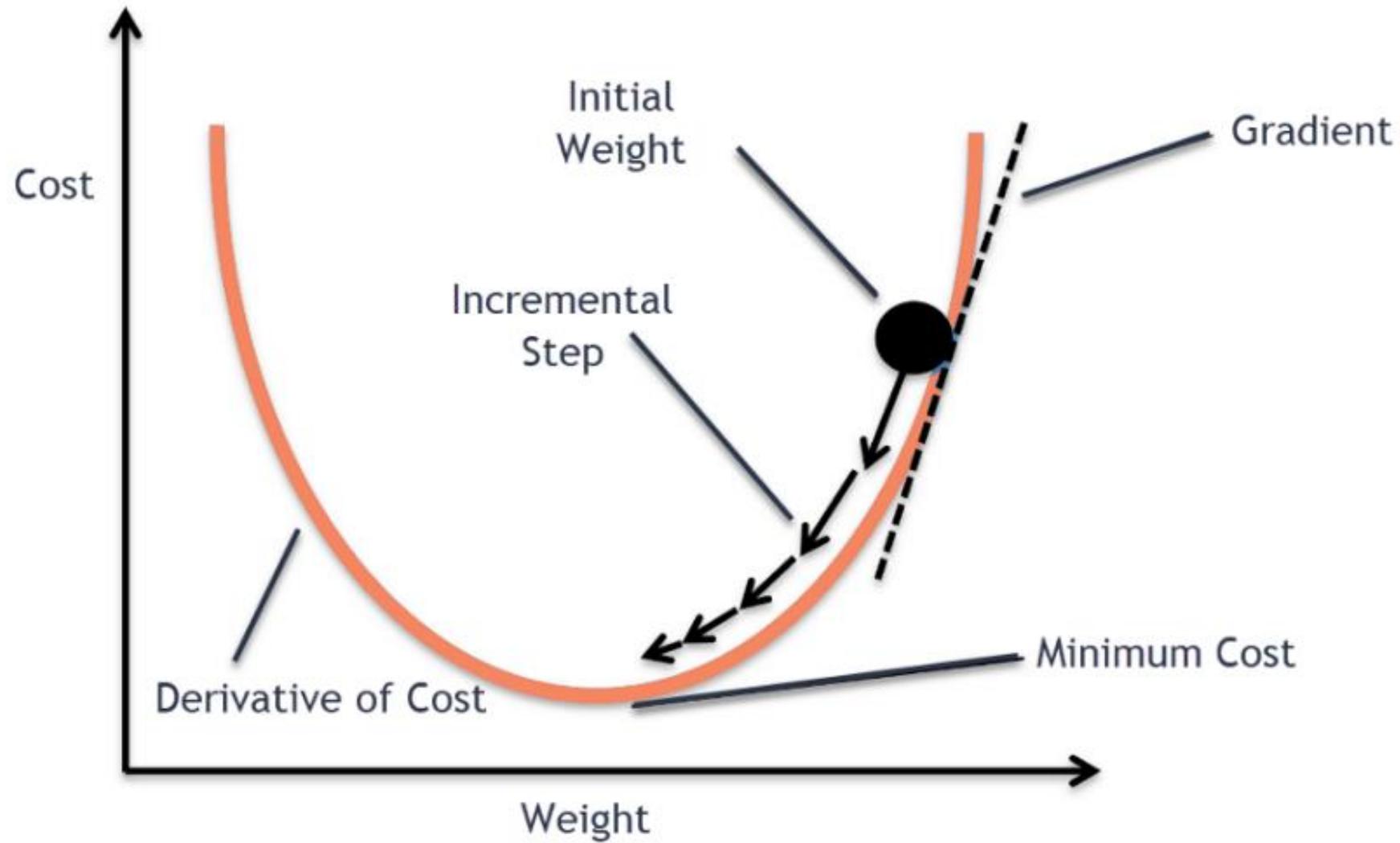
$$\frac{\partial Z}{\partial X} = \frac{\partial Z}{\partial Y} \cdot \frac{\partial Y}{\partial X}$$

Backpropagation

$$\frac{\partial \text{MSE}}{\partial W_1}(W_1) = \frac{\partial \text{MSE}}{\partial O_{out}} \frac{\partial O_{out}}{\partial O_{in}} \frac{\partial O_{in}}{\partial H_{out}} \frac{\partial H_{out}}{\partial H_{in}} \frac{\partial H_{in}}{\partial W_1}$$

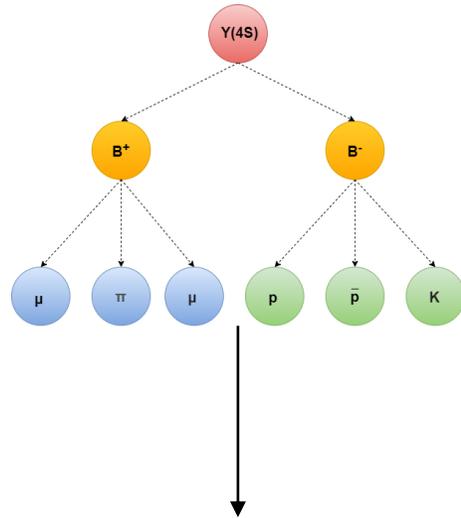
$$\frac{\partial \text{MSE}}{\partial W_2}(W_2) = \frac{\partial \text{MSE}}{\partial O_{out}} \frac{\partial O_{out}}{\partial O_{in}} \frac{\partial O_{in}}{\partial W_2}$$

Réseau de neurones

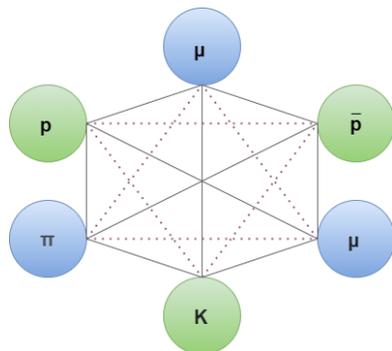


Réseau de neurones à graphes

Arbre de désintégration de $\Upsilon(4S)$



Graphe entièrement connecté entre les particules de l'état final



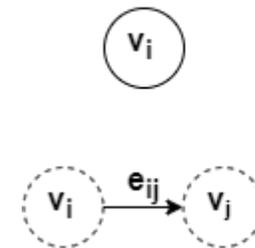
Matrice des quadri-impulsions

	P_x	P_y	P_z	E
a	a_0	a_1	a_2	a_3
b	b_0	b_1	b_2	b_3
c	c_0	c_1	c_2	c_3
d	d_0	d_1	d_2	d_3
e	e_0	e_1	e_2	e_3
f	f_0	f_1	f_2	f_3

Matrice des relations

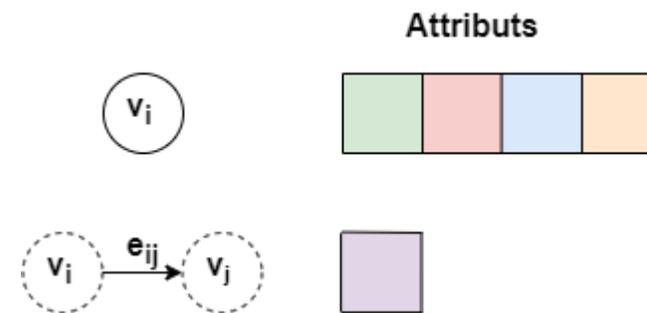
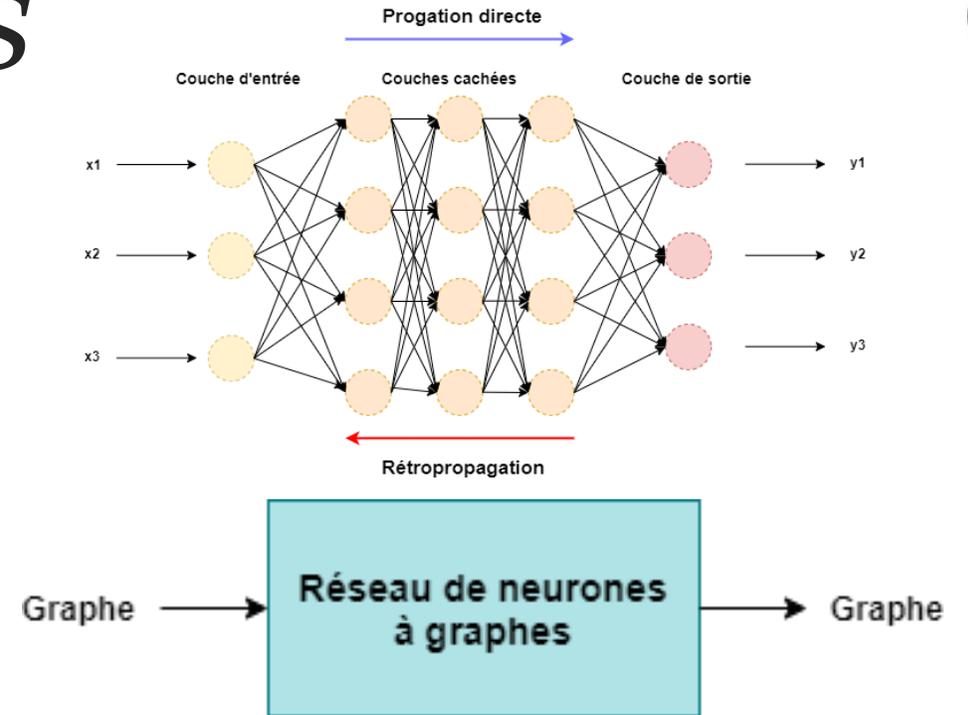
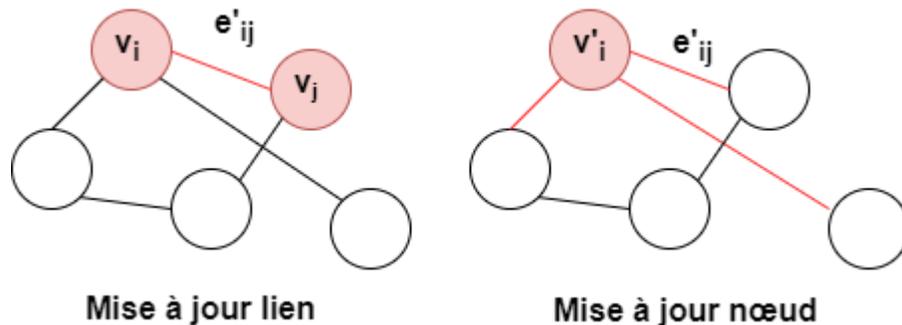
	a	b	c	d	e	f
a	0	1	1	2	2	2
b	1	0	1	2	2	2
c	1	1	0	2	2	2
d	2	2	2	0	1	1
e	2	2	2	1	0	1
f	2	2	2	1	1	0

Attributs

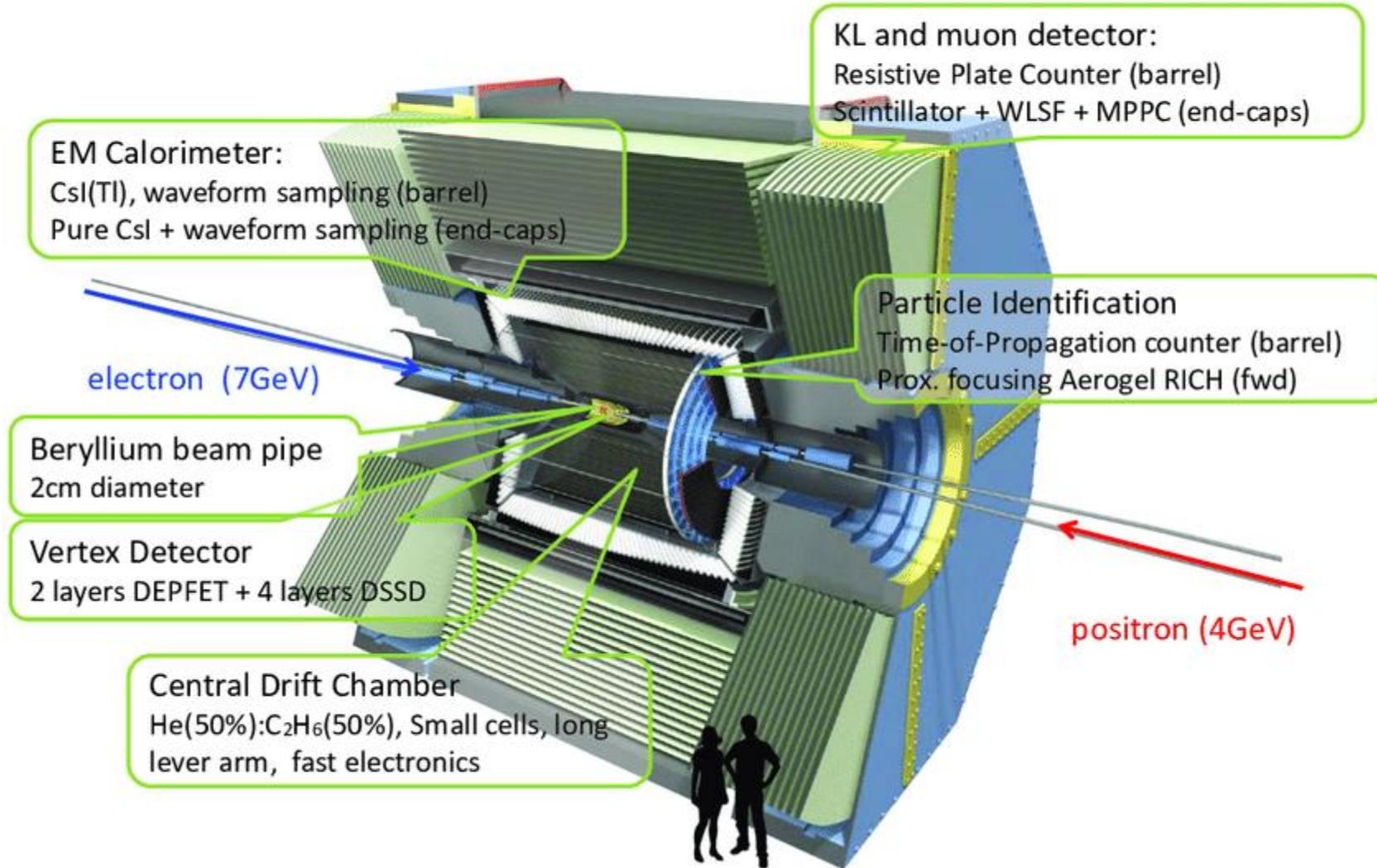


Réseau de neurones à graphes

- Apprentissage profond
- Adapté pour des données structurées
- Composé de nœuds v_i et de liens e_{ij} reliant ces nœuds
- Mise à jour du réseau :
 - Mise à jour des attributs des liens
 - Agrégation des attributs des liens sur chaque nœud
 - Mise à jour des attributs des nœuds
- $q^2=5616 \text{ MeV}^2$



Belle II Detector



Annexes

Performance 3o3

Masse racine	Masse droite	Masse gauche	q ²	Performance Validation
100	30	30	8200	92.1%
100	31	31	8078	89.0%
100	32	32	7952	86.2%
100	36	36	7408	88.0%
100	38	38	7112	82.7%
100	40	40	6800	77.8%
100	40.5	40.5	6719.5	78.4%
100	40.8	40.8	6670.72	78.0%
100	41	41	6638	69.9%
100	41.2	41.2	6605.12	72.0%
100	41.5	41.5	6555.5	69.6%
100	41.9	41.9	6488.78	70.2%
100	42	42	6472	67.0%
100	42.5	42.5	6387.5	62.6%
100	43	43	6302	72.5%
100	45	45	5950	74.0%
100	48	48	5392	75.7%
100	49	49	5198	74.1%
149	50	50	5000	91.4%

Performance 4o4

Masse racine	Masse droite	Masse gauche	q ²	Performance Validation
100	27	27	8542	99.9%
100	32	32	7952	92.4%
100	34	34	7688	83.0%
100	36	36	7408	70.9%
100	38	38	7112	68.1%
100	40	40	6800	39.2%
100	40.5	40.5	6719.5	41.5%
100	41	41	6638	34.4%
100	41.5	41.5	6555.5	30.1%
100	42	42	6472	7.43%
100	43	43	6302	1.95%
100	44	44	6128	2.64%
100	45	45	5950	4.47%
100	46	46	5768	0.94%
100	47	47	5582	1.55%
100	48	48	5392	1.74%
100	49	49	5198	7.17%
100	49.5	49.5	5099.5	46.5%
149	50	50	5000	68.4%

Annexes

Performance 5o5

Masse racine	Masse droite	Masse gauche	q ²	Performance Validation
100	30	30	8200	99.9%
100	32	32	7952	98.9%
100	34	34	7688	94.5%
100	36	36	7408	84.7%
100	38	38	7112	58.5%
100	39	39	6958	53.3%
100	40	40	6800	53.6%
100	41	41	6638	41.6%
100	42	42	6472	19.7%
100	42.5	42.5	6387.5	13.9%
100	43	43	6302	1.97%
100	44	44	6128	0.43%
100	45	45	5950	0.31%
100	48	48	5392	0.01%
100	48.5	48.5	5295.5	0.06%
100	49	49	5198	0.11%
100	49.5	49.5	5099.5	0%

Performances d'entraînement et de validation pour différentes topologies.

Nombre FSP	Topologie	Performance Training	Performance Validation
4	2o2	99.1%	99.1%
5	3o2	98.6%	98.5%
6	3o3	91.8%	90.3%
6	4o2	99.8%	99.8%
7	4o3	74.7%	58.7%
7	5o2	95.5%	95.5%
8	3o5	84.3%	72.3%
8	4o4	34.9%	29.5%
8	6o2	100%	99.8%
9	4o5	4.04%	3.47%
9	6o3	67.8%	67.0%
9	7o2	100%	99.9%
10	3o7	63.9%	59.1%
10	6o4	6.38%	5.54%
10	5o5	1.70%	1.65%
10	8o2	100%	99.9%

Annexes

Distribution des quadri-impulsions : exemples

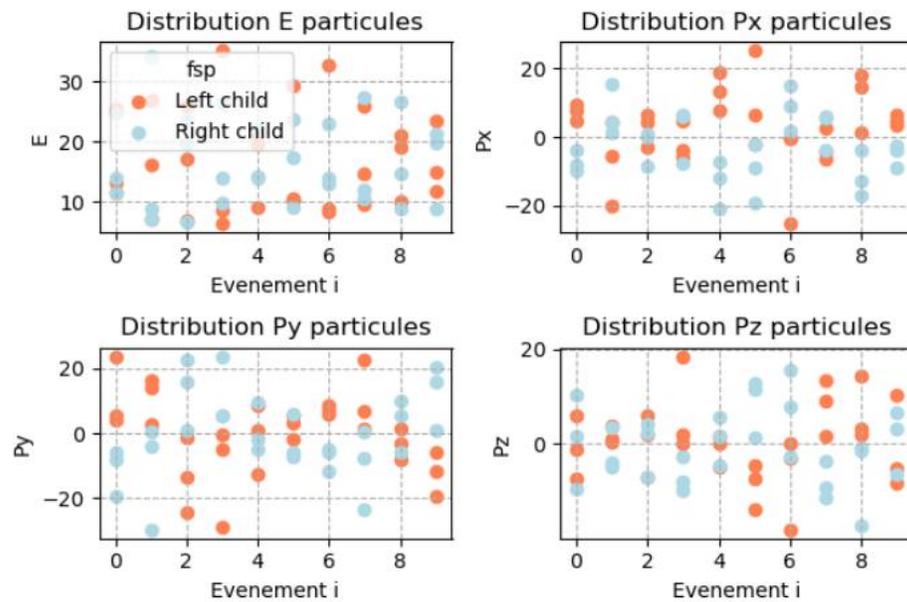


FIGURE F.1 – Distribution de la quadri-impulsion pour $q^2=8200 \text{ MeV}^2$ pour un échantillon de dix évènements. FSP : six particules. Topologie 3o3.

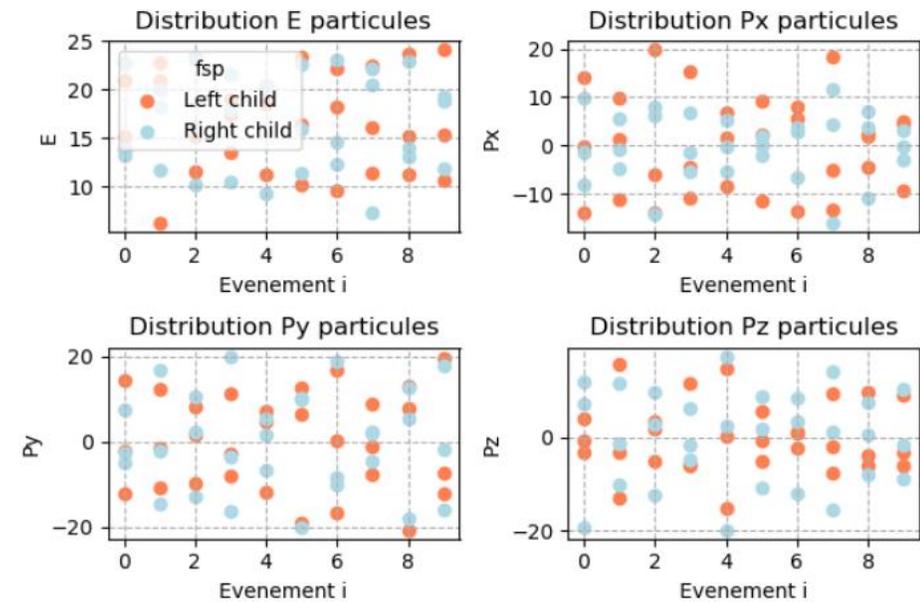


FIGURE F.2 – Distribution de la quadri-impulsion pour $q^2=5000 \text{ MeV}^2$ pour un échantillon de dix évènements. FSP : six particules. Topologie 3o3.

Annexes

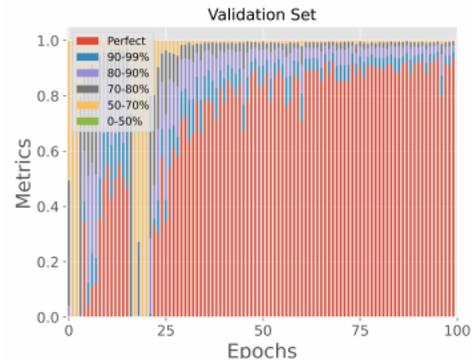


FIGURE D.1 – Performance de validation d'un groupe de données de topologie 3o3. FSP : six particules de l'état final.

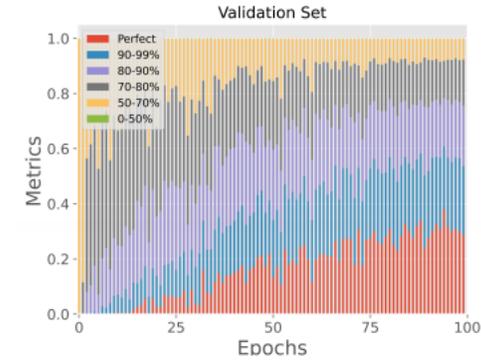


FIGURE D.3 – Performance de validation d'un groupe de données de topologie 4o4. FSP : huit particules de l'état final.

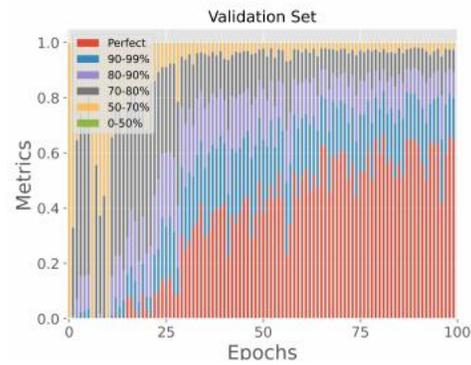


FIGURE D.2 – Performance de validation d'un groupe de données de topologie 4o3. FSP : sept particules de l'état final.

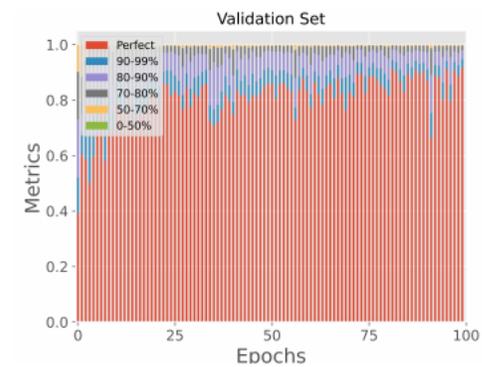


FIGURE D.4 – Performance de validation d'un groupe de données. FSP : entre quatre et six. 2500 évènements par topologie.

Annexes

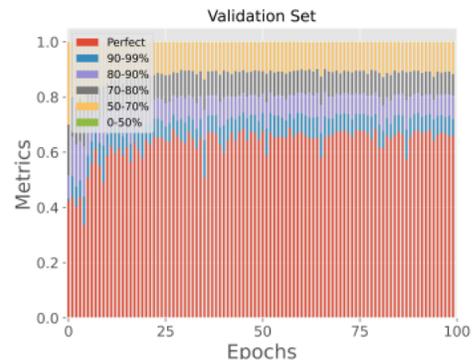


FIGURE D.5 – Performance de validation d'un groupe de données. FSP : entre quatre et dix. 2500 évènements par topologie.