



Istituto Nazionale di Fisica Nucleare

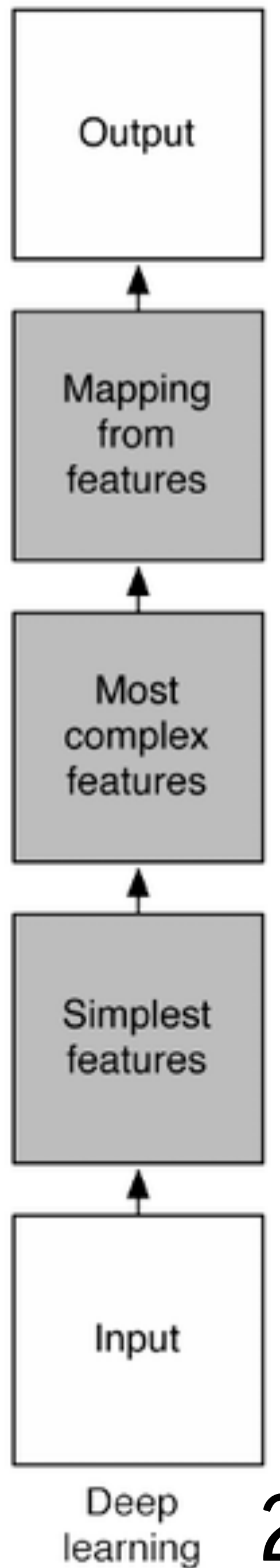
# AI & MACHINE LEARNING

Stefano G.

Retreat Fisica Particelle Elementari - INFN Roma - 16-18.6.2019 - Assisi

# ML: CHE ROBA È?

- nella accezione più comune il ML è un insieme di metodi computazionali in grado di imparare dai dati
  - costruire rappresentazioni efficaci dei dati analizzati (deep representation learning)
  - apprendere il set di regole (relazioni funzionali) che lega tra loro rappresentazioni dei dati e risultati di un dato compito
- basandosi unicamente su campioni di esempi e su una misura delle prestazioni ...



# ML E NN: UNIVERSAL APPROXIMATION THEOREM

Let  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  be a nonconstant, **bounded**, and **continuous** function. Let  $I_m$  denote the  $m$ -dimensional **unit hypercube**  $[0, 1]^m$ . The space of real-valued continuous functions on  $I_m$  is denoted by  $C(I_m)$ . Then, given any  $\varepsilon > 0$  and any function  $f \in C(I_m)$ , there exist an integer  $N$ , real constants  $v_i, b_i \in \mathbb{R}$  and real vectors  $w_i \in \mathbb{R}^m$  for  $i = 1, \dots, N$ , such that we may define:

$$F(x) = \sum_{i=1}^N v_i \varphi(w_i^T x + b_i)$$

as an approximate realization of the function  $f$ ; that is,

$$|F(x) - f(x)| < \varepsilon$$

for all  $x \in I_m$ . In other words, functions of the form  $F(x)$  are **dense** in  $C(I_m)$ .

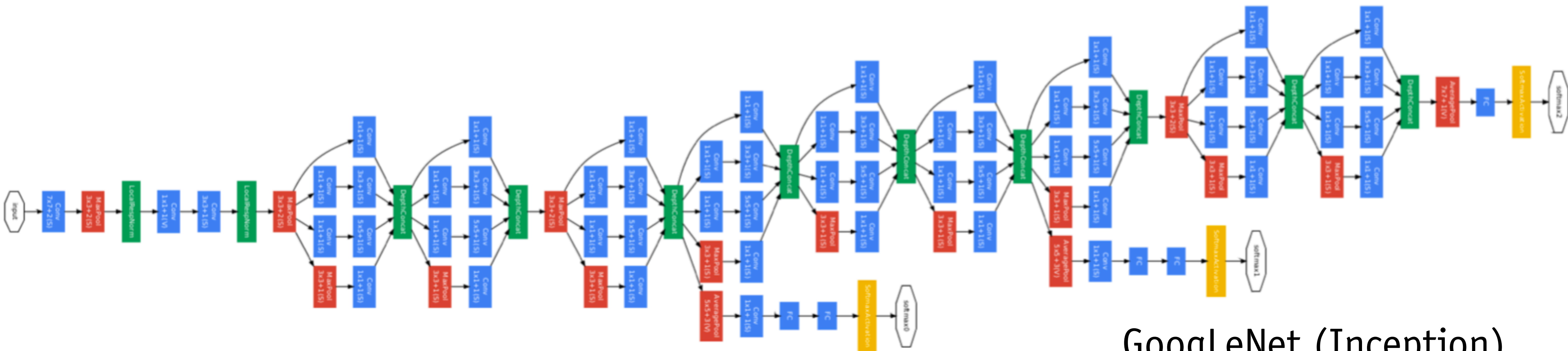
Struttura	Regioni di decisione	Forma generale
	Semispazi delimitati da iperpiani	
	Regioni convesse	
	Regioni di forma arbitraria	

**NOTA:** il teorema non dice nulla sull'effettiva possibilità di riuscire ad apprendere in modo semplice i parametri della rete ...

# DEEP LEARNING E DNN

nel Deep Learning il mapping viene migliorato organizzando idee e concetti in modo gerarchico (rappresentazioni complesse dei dati analizzati vengono espresse in termini di altre rappresentazioni più semplici e generali)

la versatilità delle reti neurali profonde ne facilita l'implementazione pratica

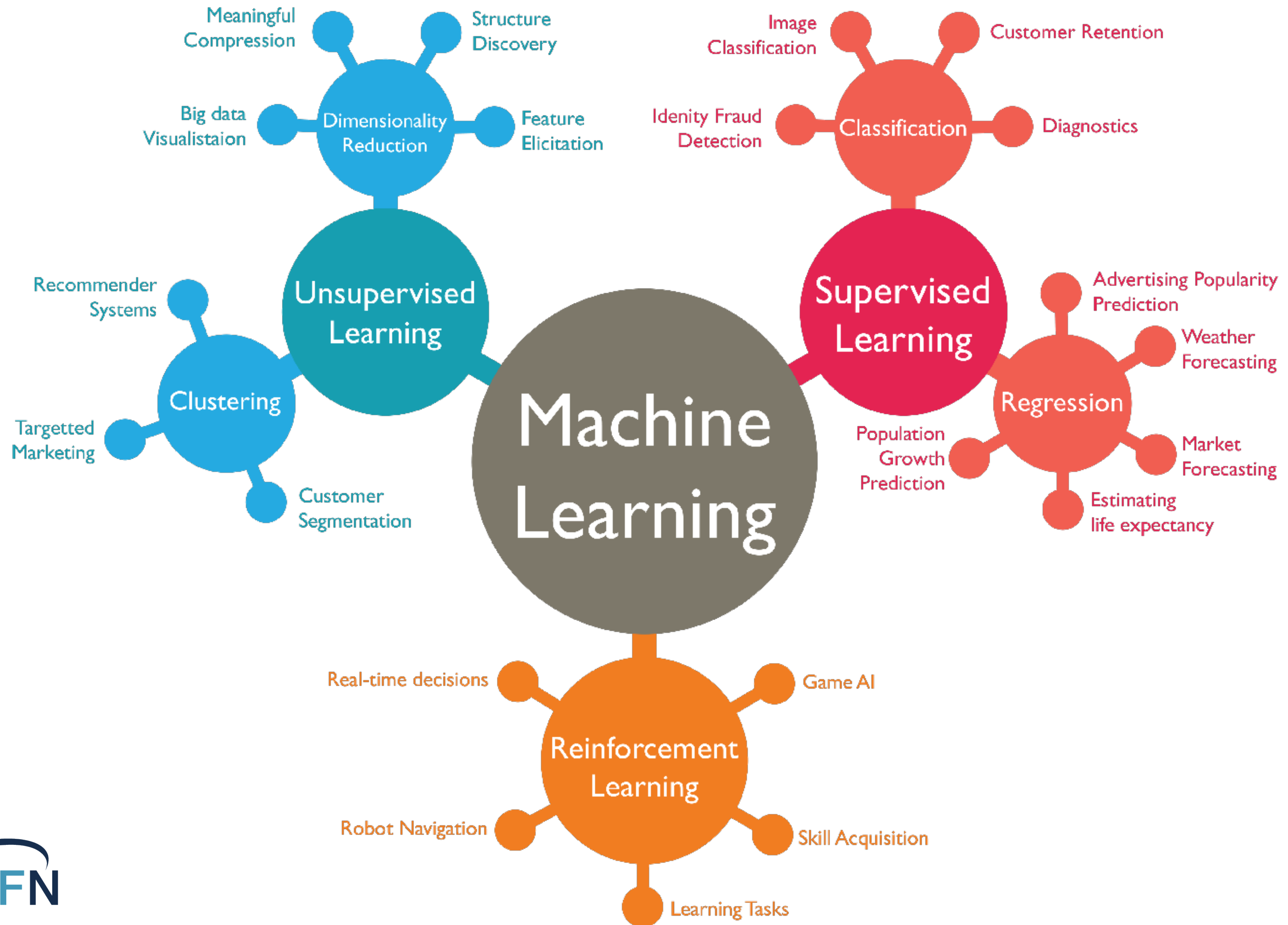


GoogLeNet (Inception)

7 MPar

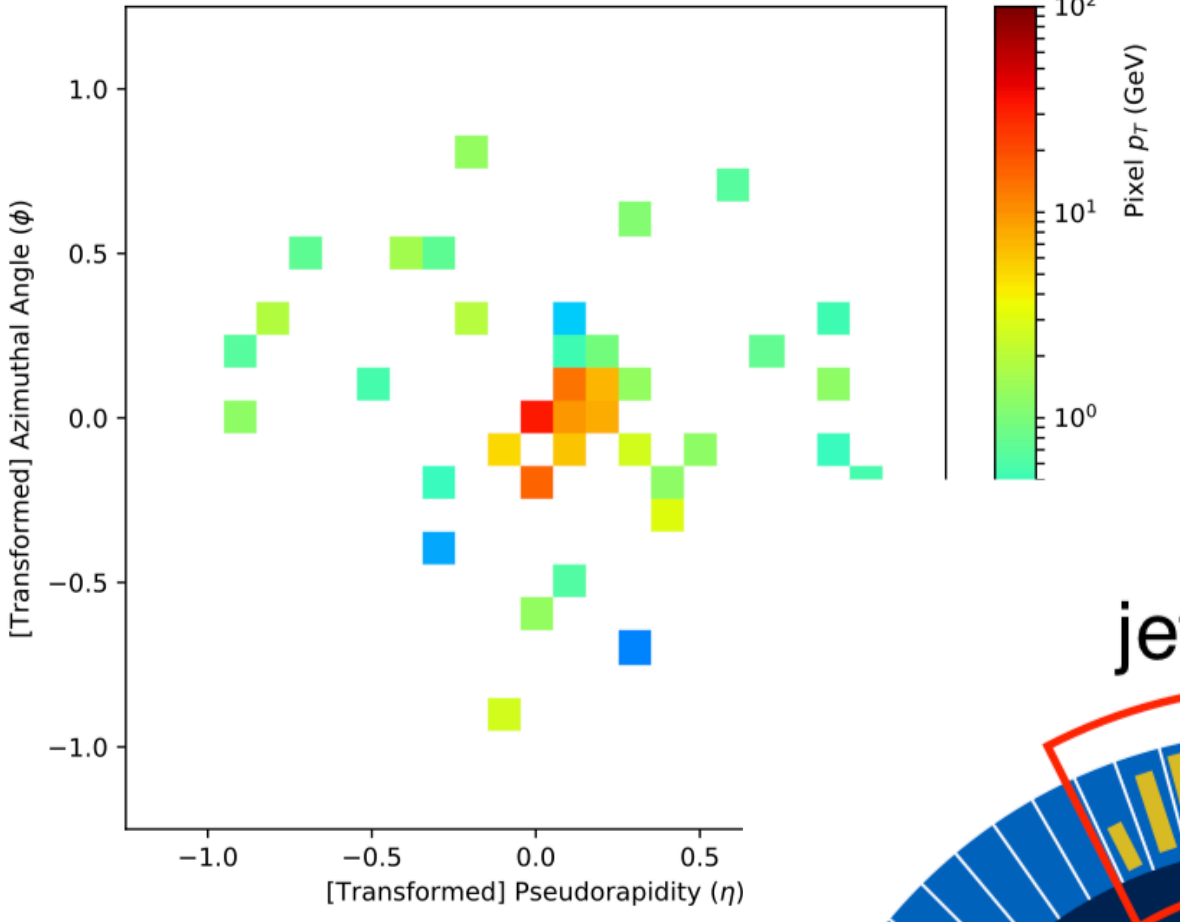
# COSA NON È / NON FA ...

- NON è il sistema definitivo per costruire sistemi AGI (i.e. in grado di sostituirci come alcuni fisici temono)
- NO FREE LUNCH THEOREM: mediato su tutte le possibili distribuzioni di probabilità dei dati di training, ogni algoritmo ML presenta lo stesso rate di errore quando applicato a dati non osservati in precedenza (i.e. funziona bene tanto quanto lanciare una moneta)
- e soprattutto non è in grado di “migliorare” i dati in input ...
- UGLY DUCKLING THEOREM: in assenza di un qualche criterio di preferenza/bias iniziale NON esiste nessuna rappresentazione delle feature privilegiata (i.e. non è possibile classificare in assenza di bias)

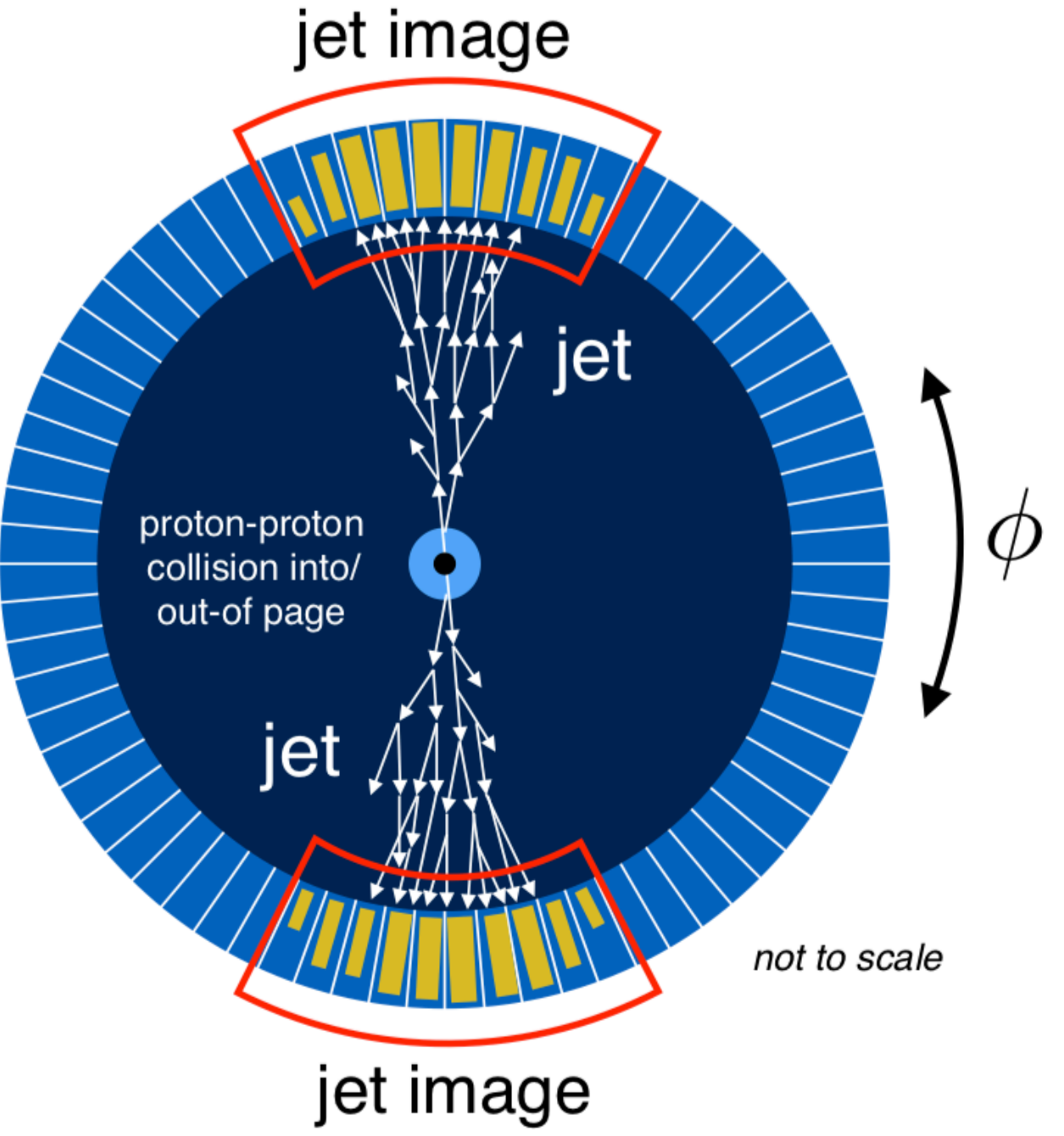


# APPLICAZIONI IN FISICA

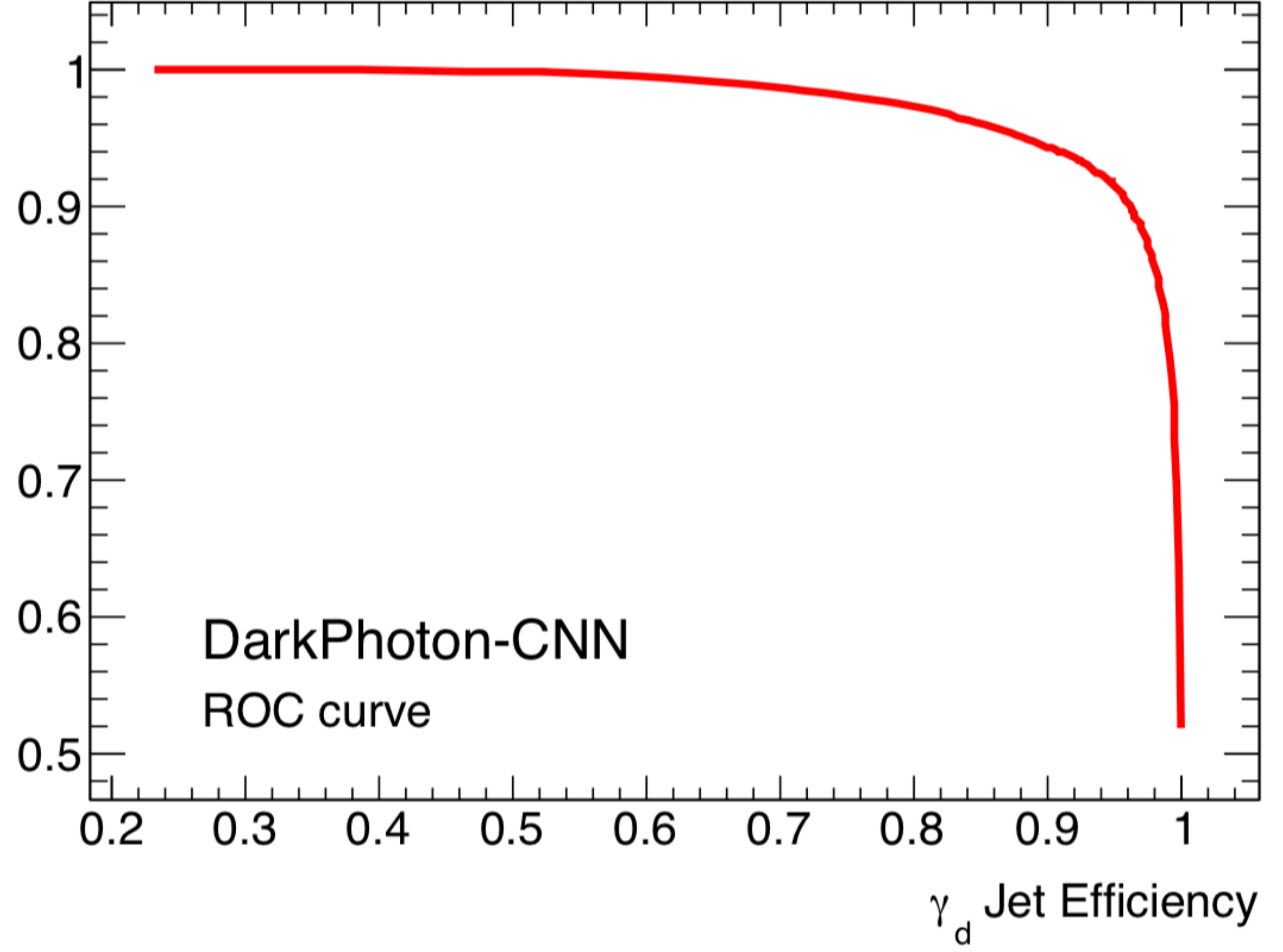
## Real image example



ConvNET per la classificazione di jet adronici anomali



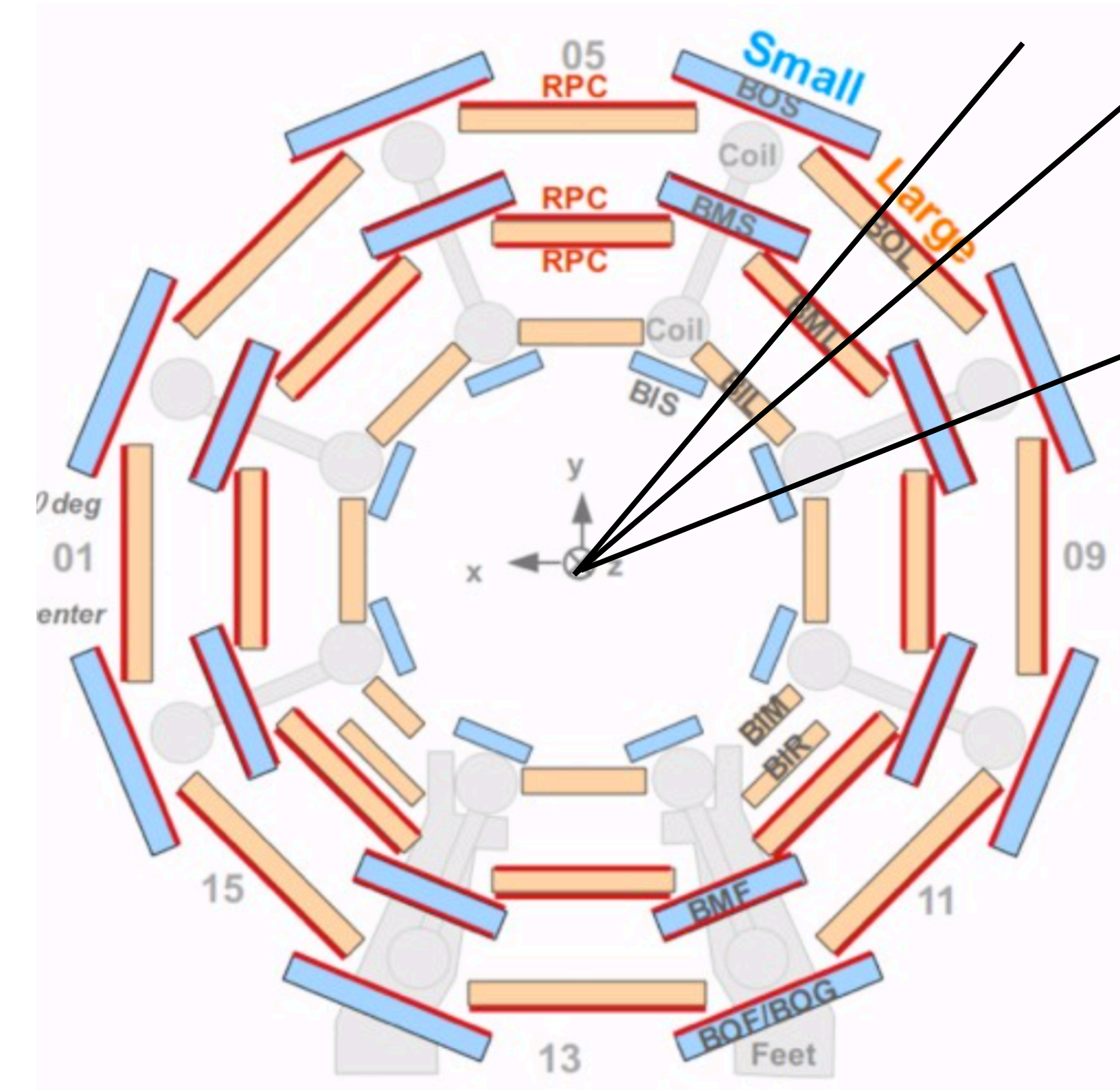
QCD Jet Rejection



$pp \rightarrow qq \rightarrow 2 \text{ jet}$   
VS

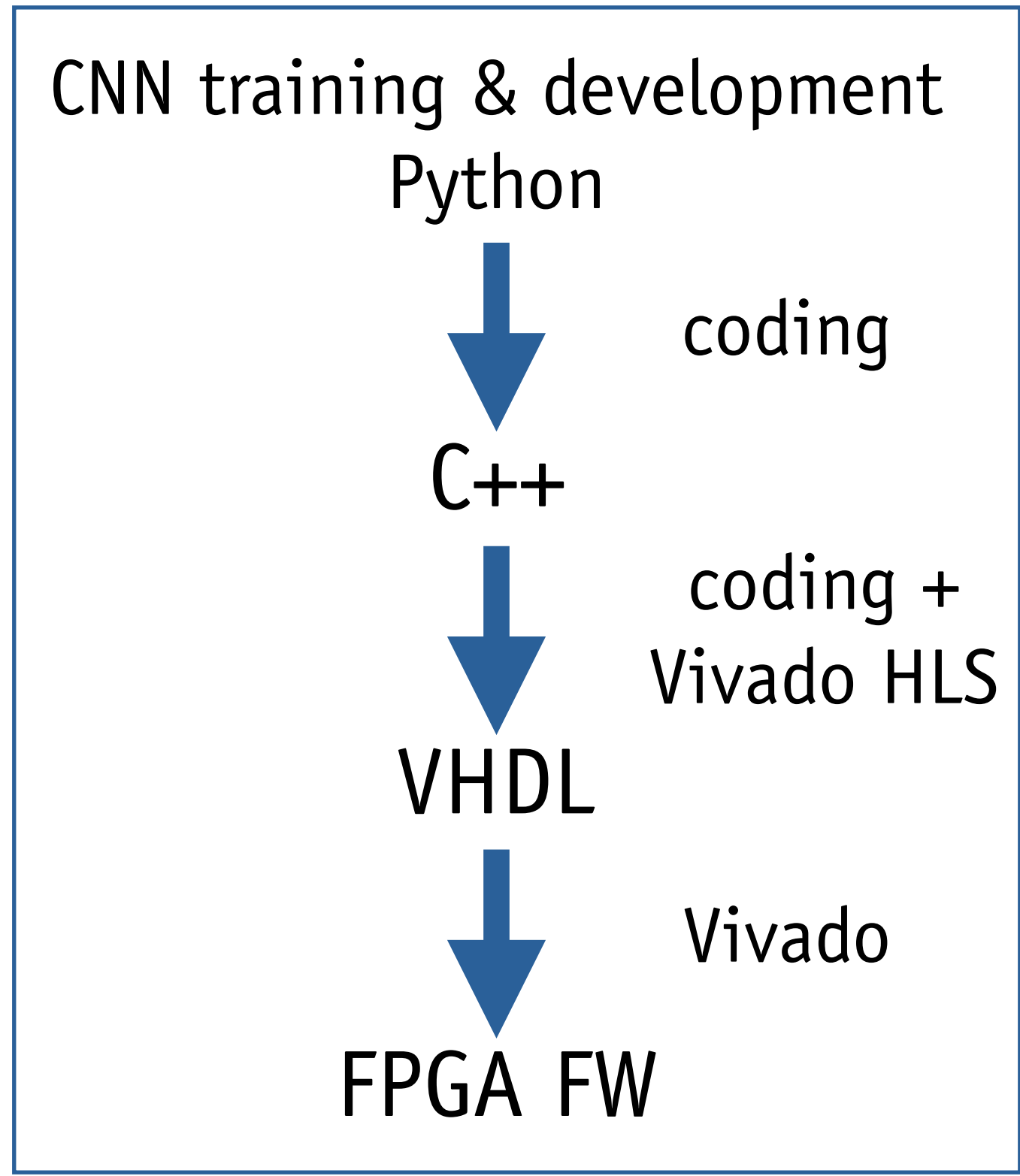
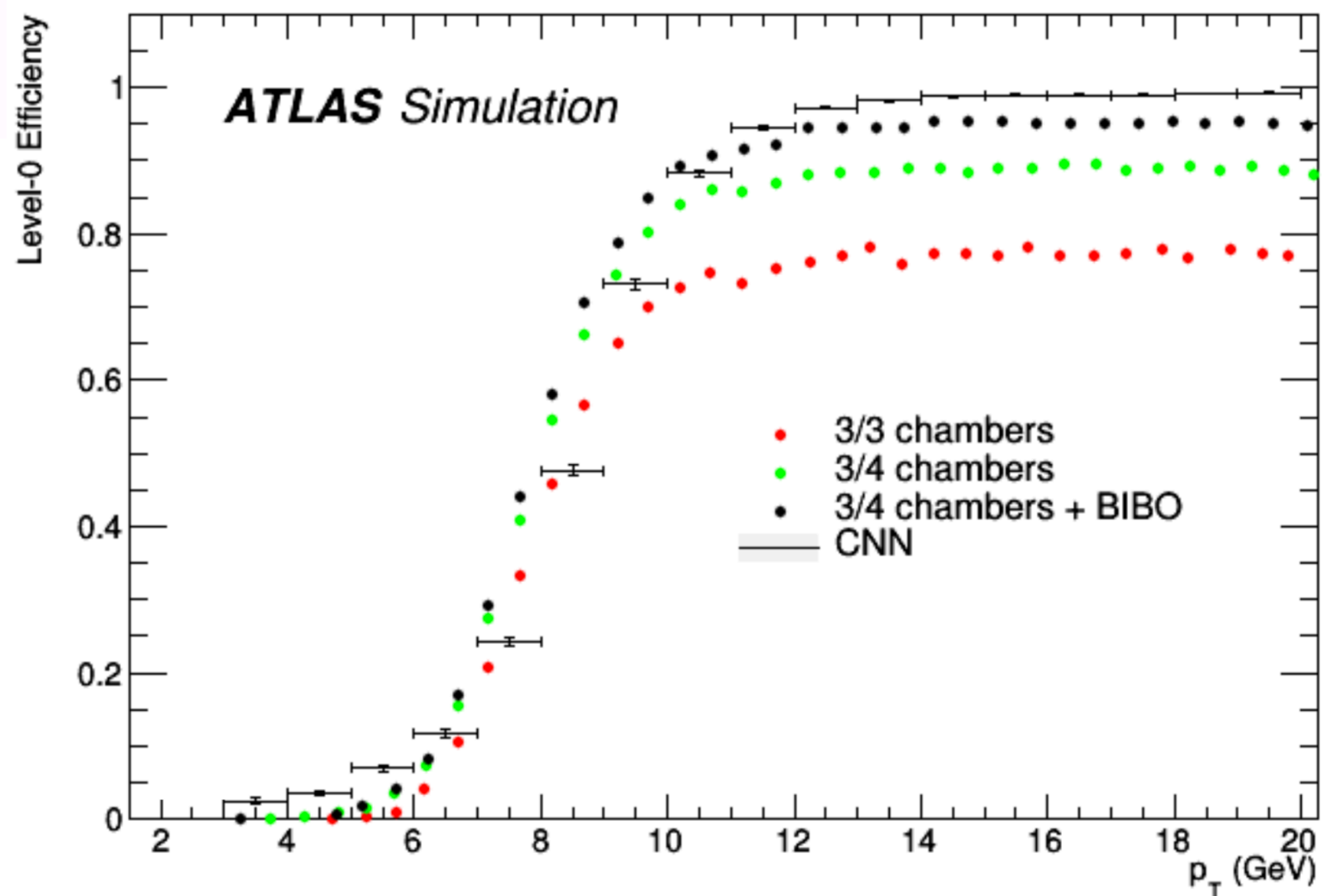
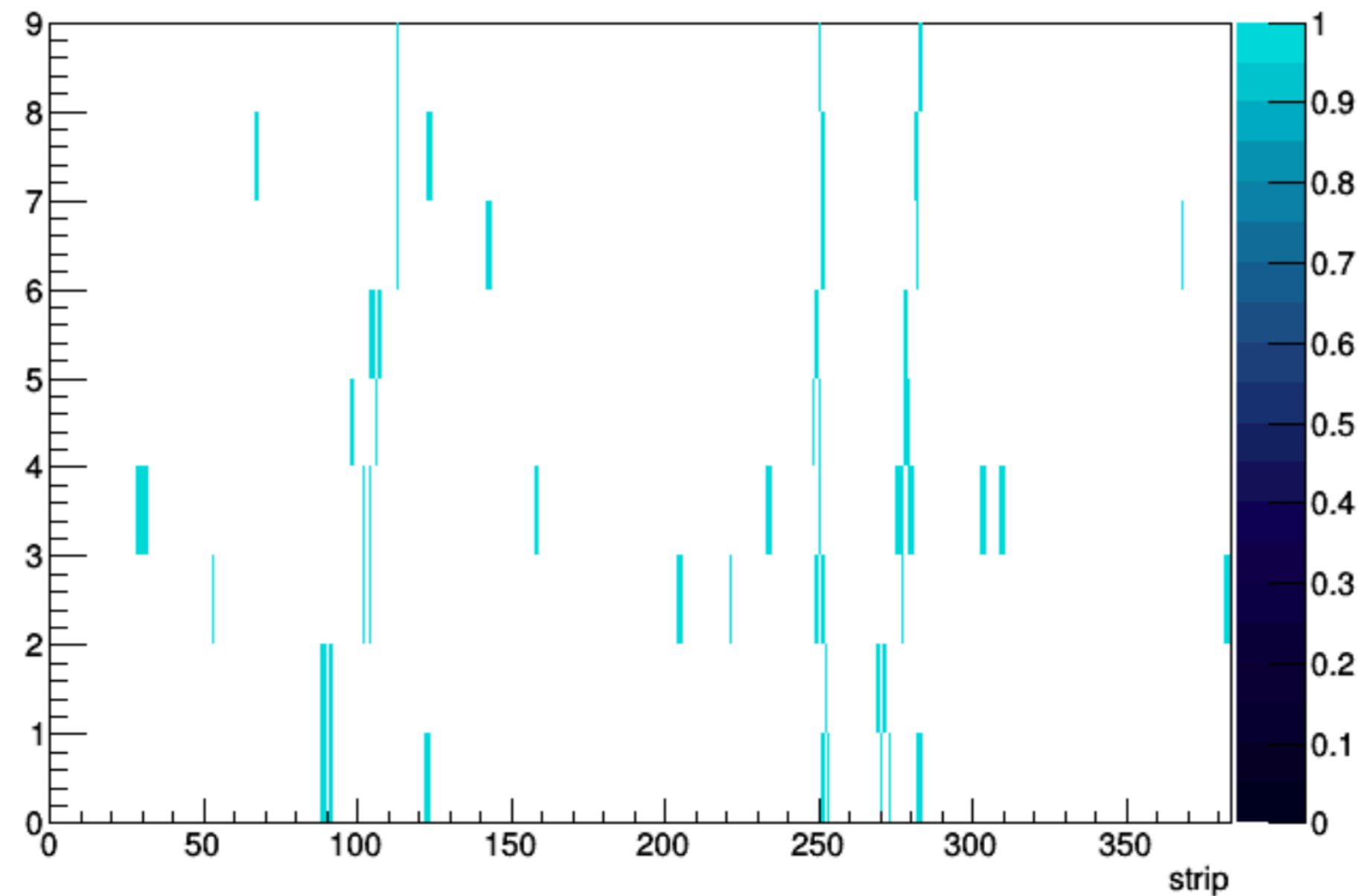
$pp \rightarrow \Upsilon_d \Upsilon_d \rightarrow \text{displaced dark-photons} \rightarrow 2 \text{ displaced jet}$  7





Ricostruzione ultraveloce di muoni con CNN-ternarie su FPGA per il trigger di L0 di ATLAS FASE-2

3 Muons + background

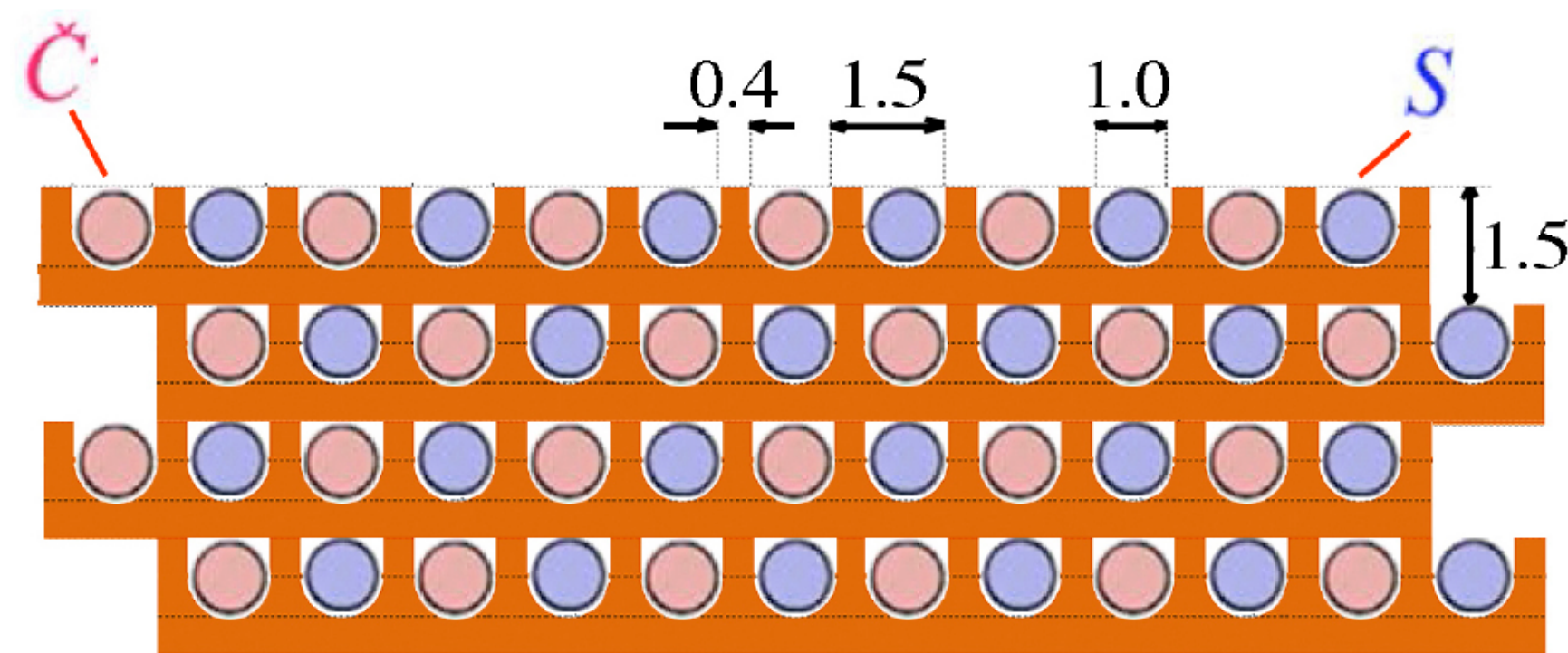


target tempo processamento  $\approx 2\mu s$

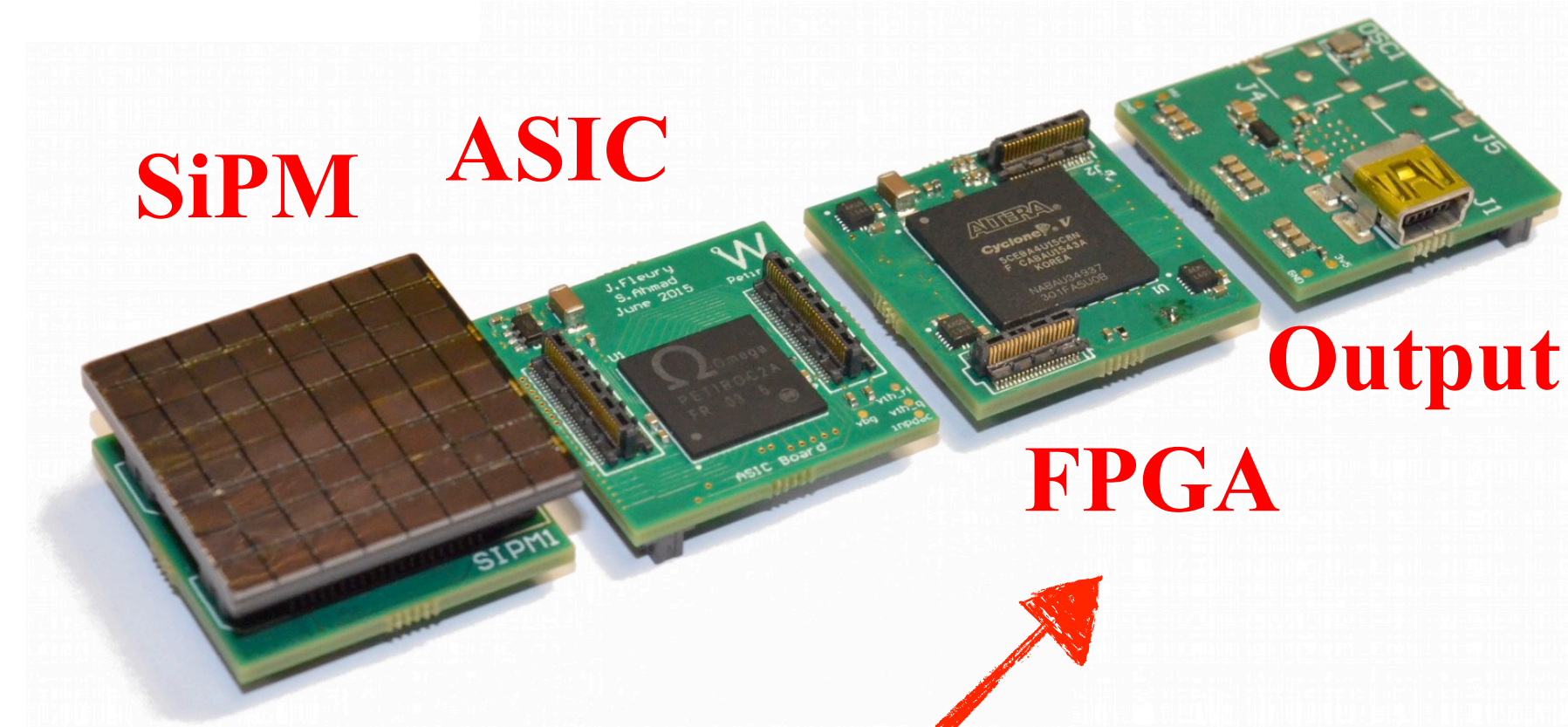
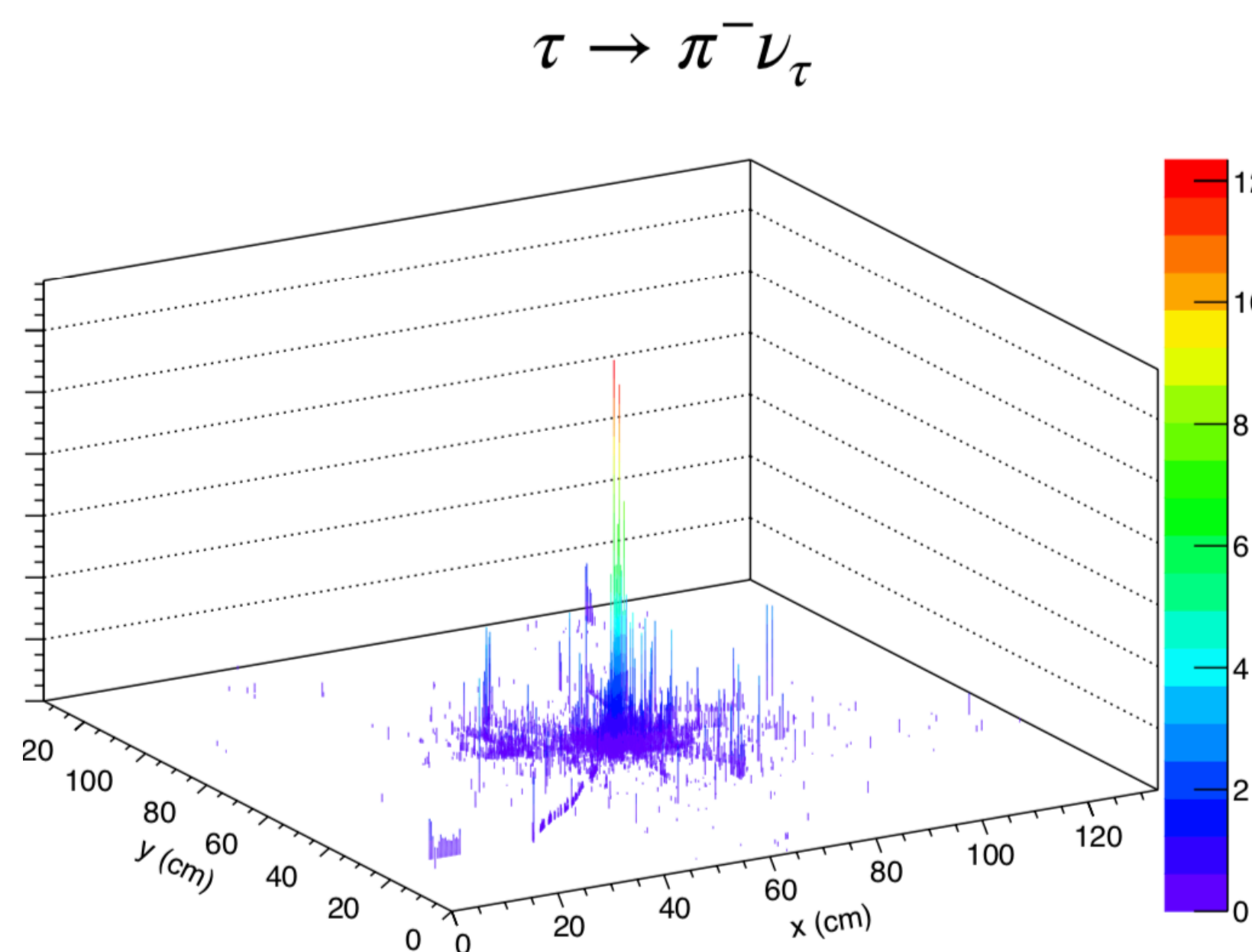
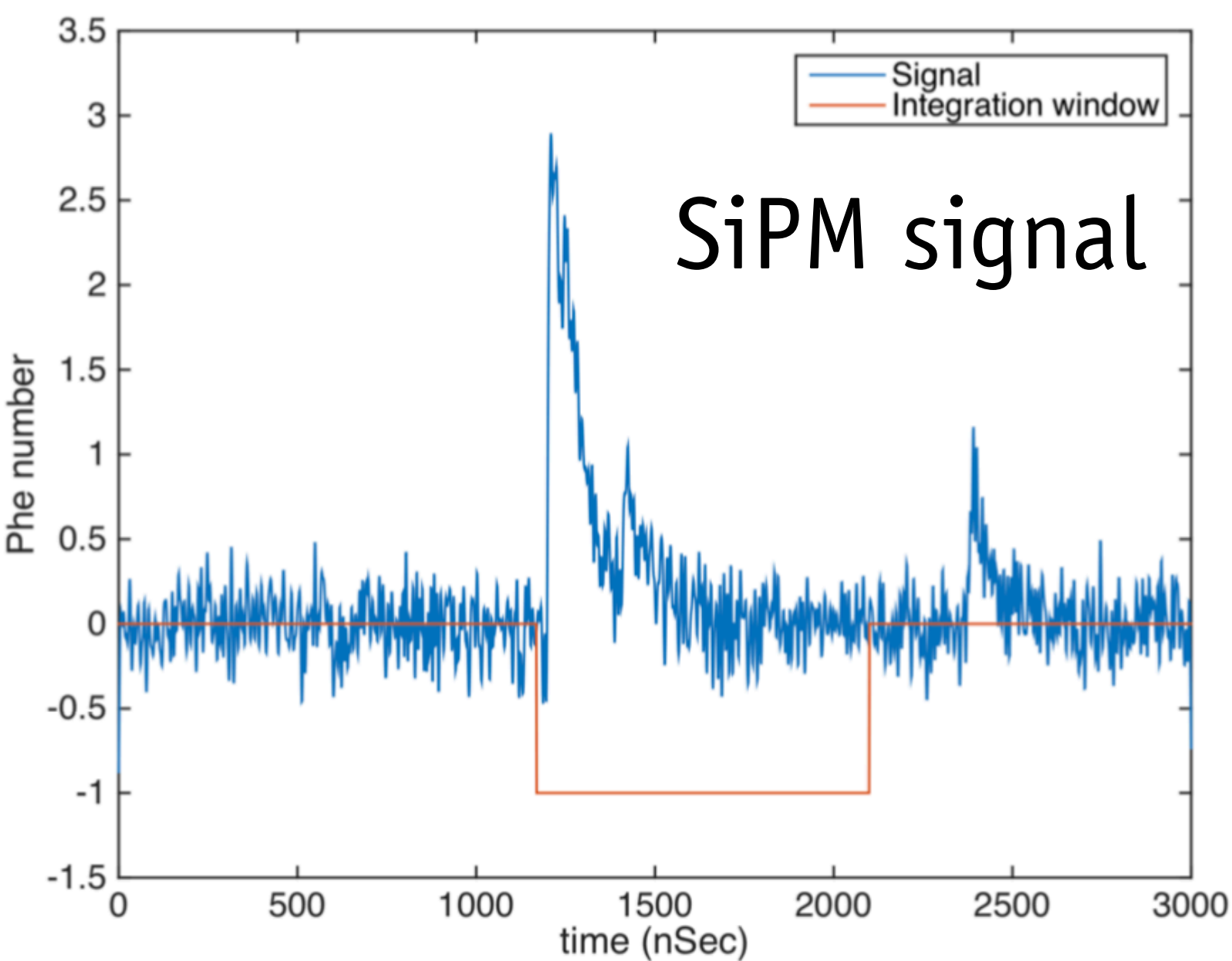


# A spin-off del lavoro fatto in ATLAS

identificazione in tempo reale di decadimenti di leptoni  $\tau$  con il calorimetro dual readout IDEA



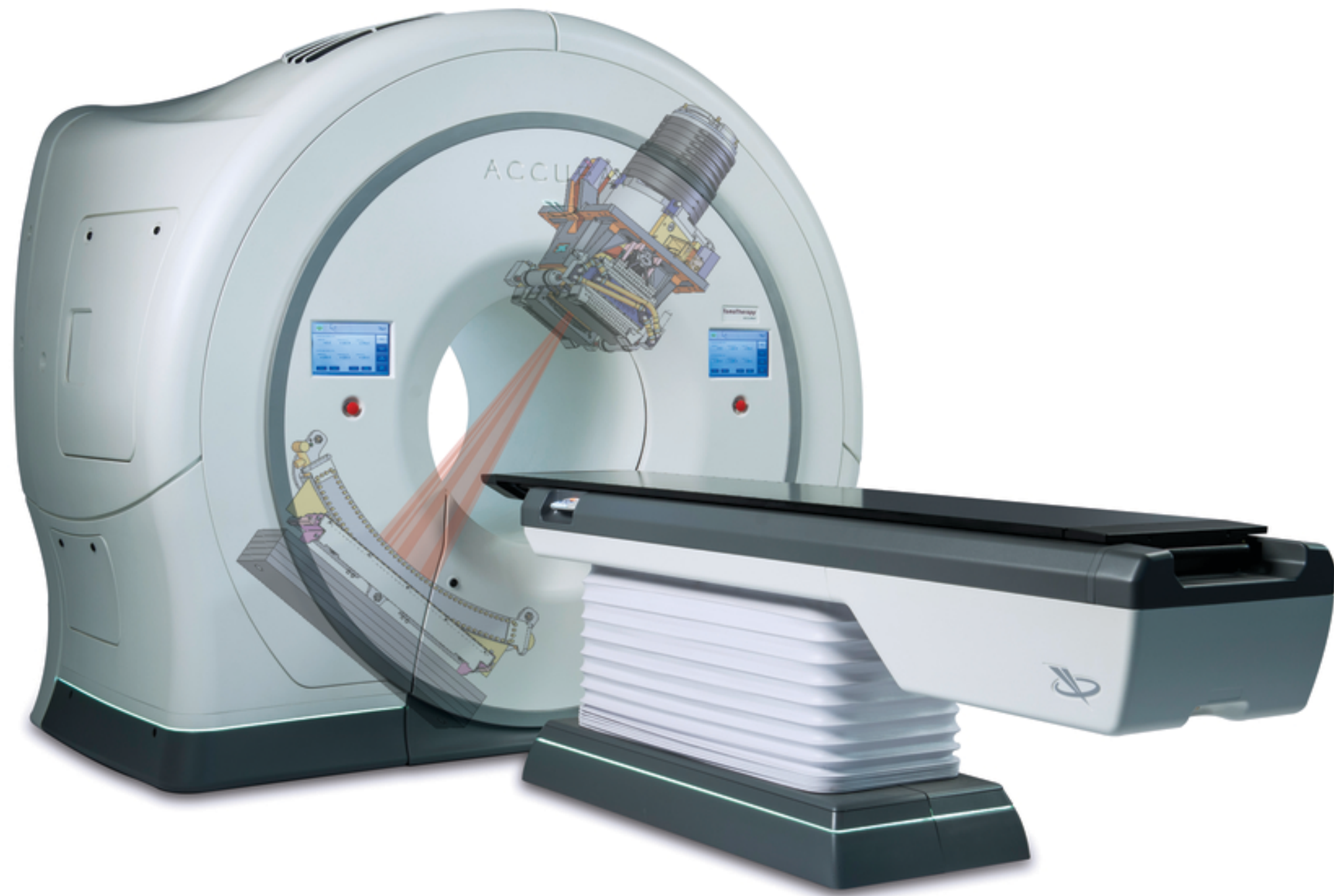
Fiber pattern RD52



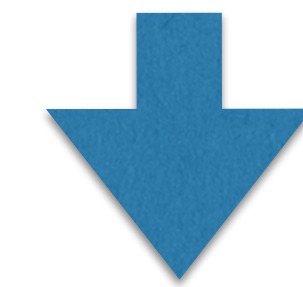
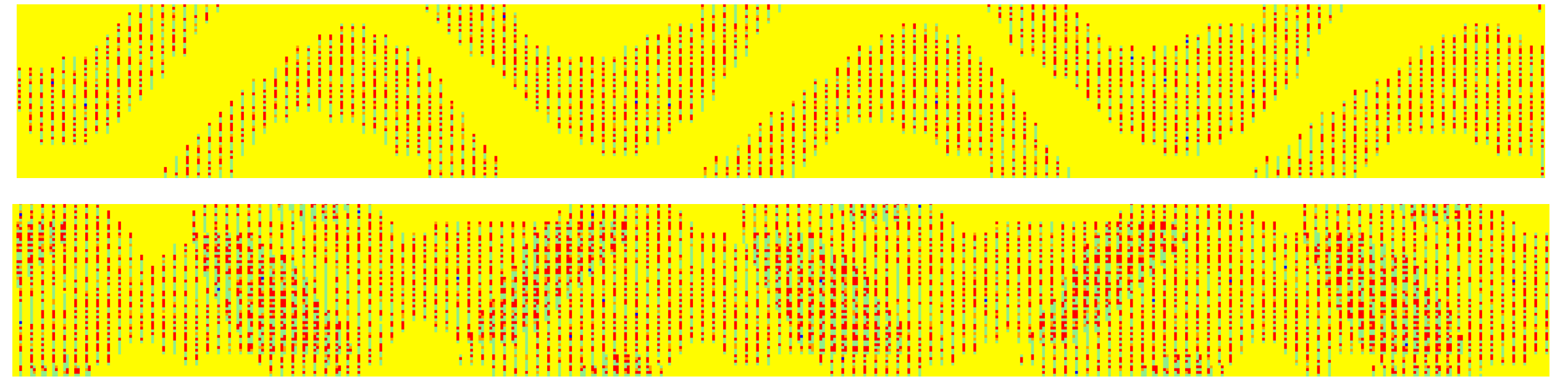
RNN  
CNN  
CNN+RNN

# DL applicato alla fisica medica

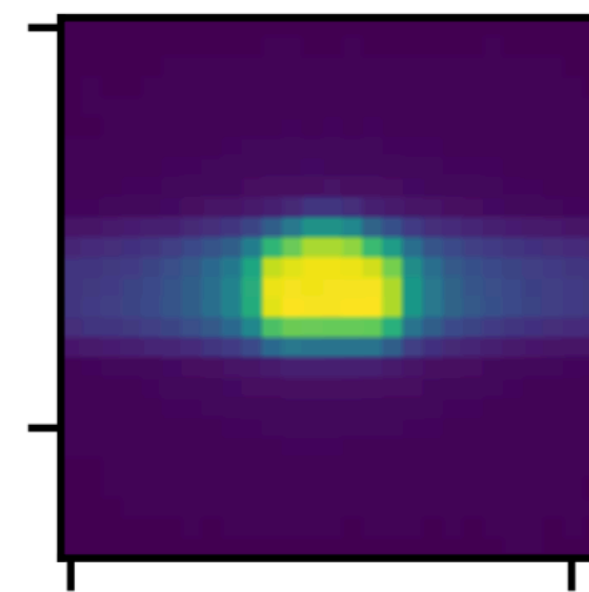
deep image to image regression con CNNs per calcolo e validazione dosi depositate in trattamenti tomoterapici



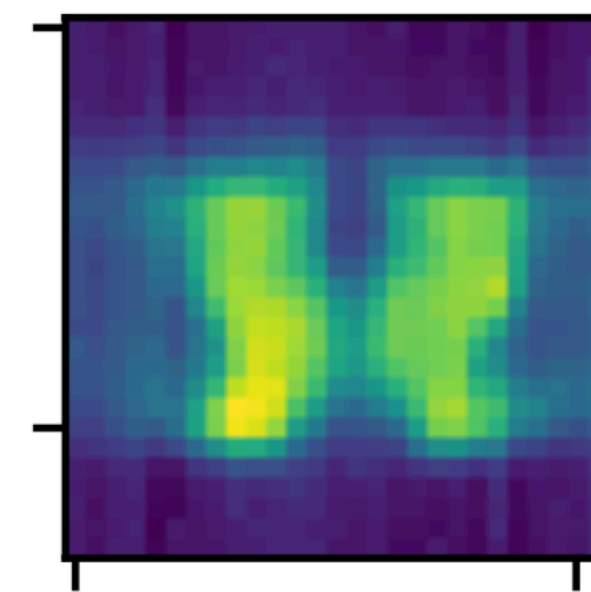
## SINOGRAMMA



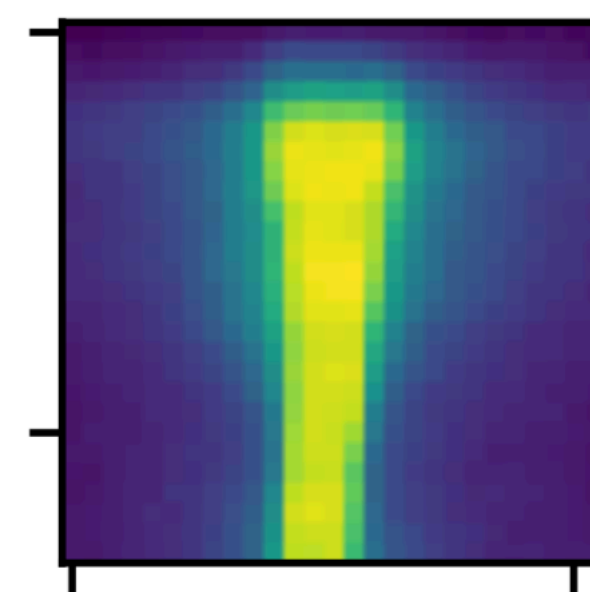
CNN+DNN → DNN



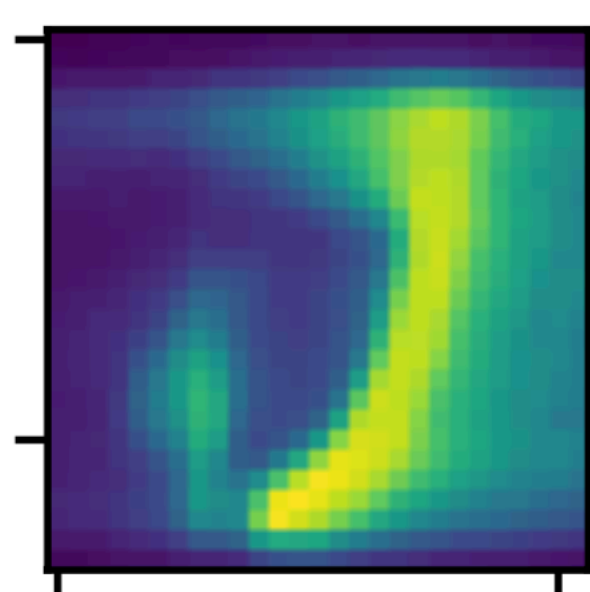
MSE: 0.000, SSIM: 0.969  
HD: 0.13, PSNR: 43.0



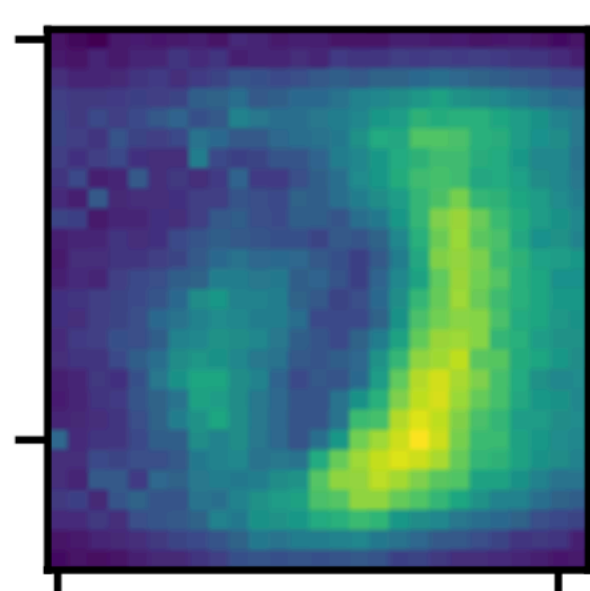
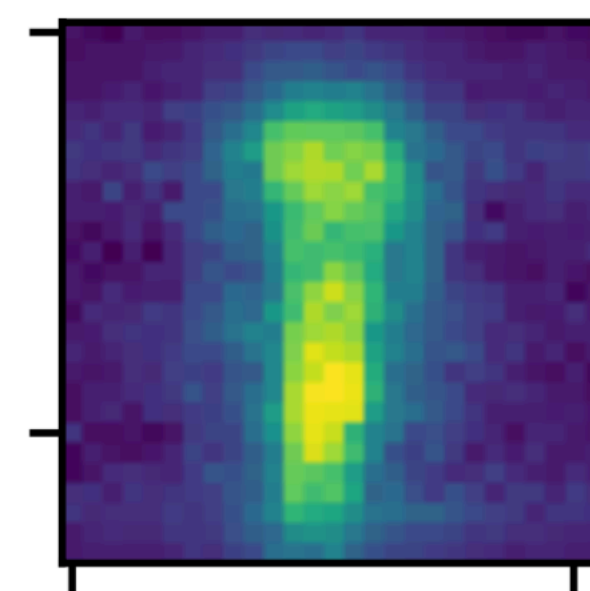
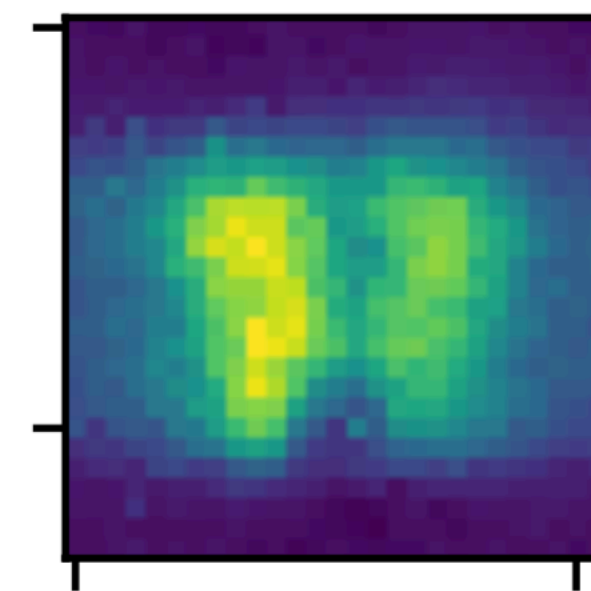
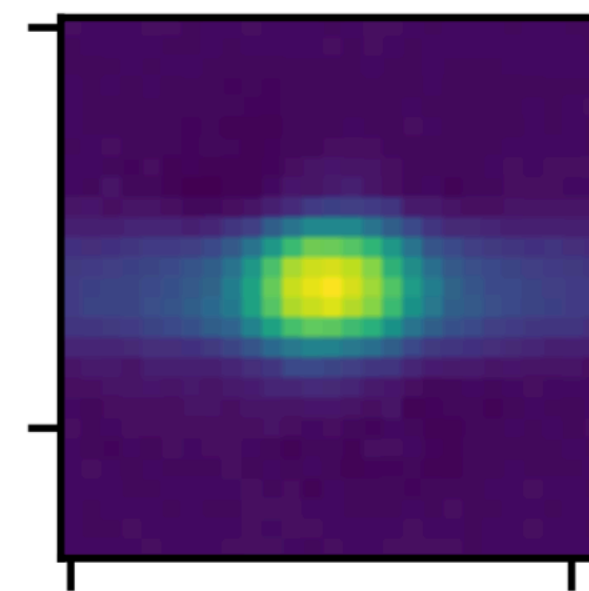
MSE: 0.000, SSIM: 0.929  
HD: 0.10, PSNR: 42.9



MSE: 0.000, SSIM: 0.937  
HD: 0.07, PSNR: 43.1

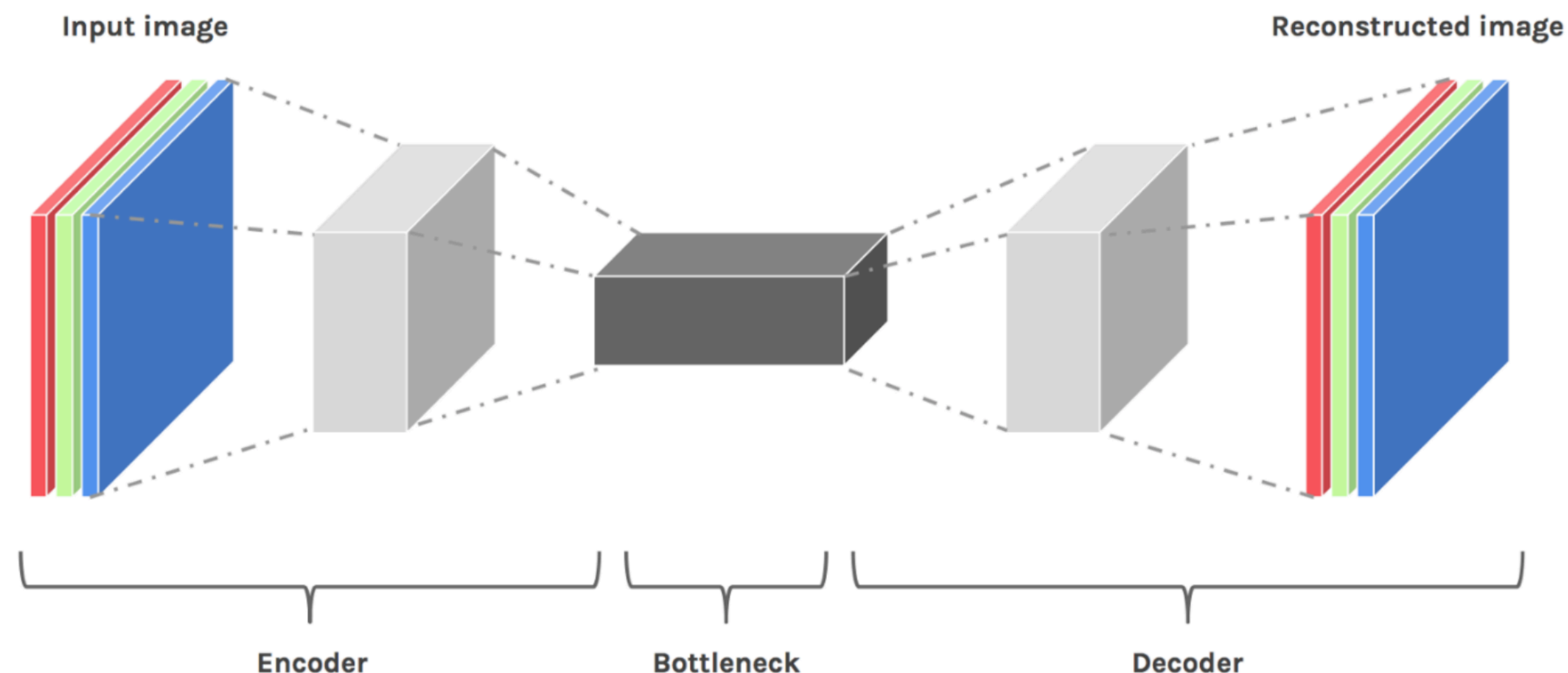


MSE: 0.001, SSIM: 0.860  
HD: 0.15, PSNR: 38.0



## Site Operation Anomaly Detection

sistema globale di rilevamento delle anomalie distribuito, basato sull'analisi dei dati ES / Beats e su DL, per monitorare le attività dei Tier-2 Italiani



Tier-2 italiani: dati auditbeat inviati a Roma(average ~ 500 Hz collected data, 10 GB/day

Anomaly detector system basato su Auto-Encoders e Variational-AE:  
transparent distributed firewall, intrusion detections, hardware failures, process misbehaviour or malicious attacks, etc

simile architettura di DNN verrà utilizzata per image-denoising nel progetto NEPTUNE

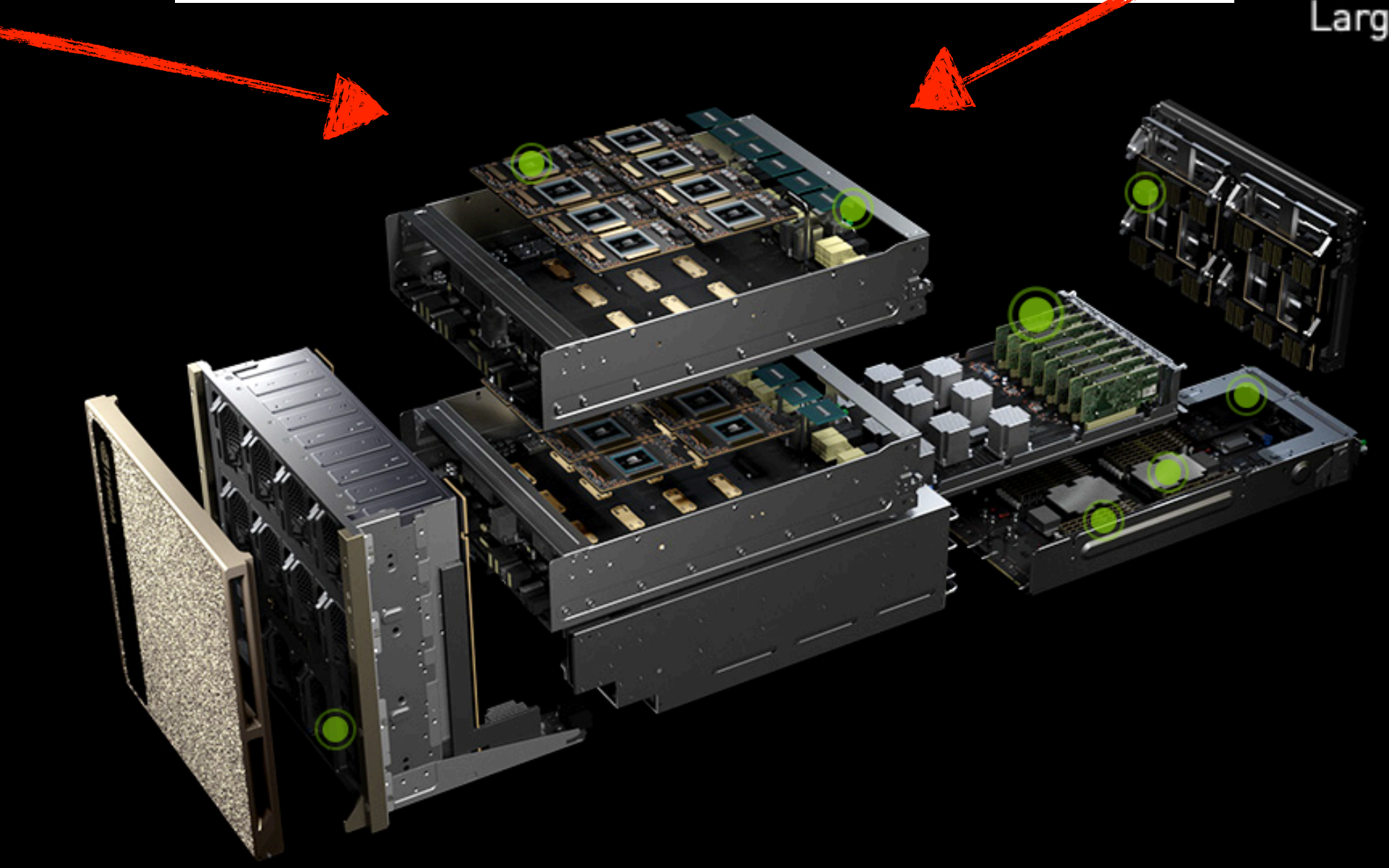
# NUOVE IDEE / SVILUPPI INTERESSANTI

- nuovi algoritmi: attention, GAN + self-attention (SAGAN), reinforcement learning, Dense Networks, ...
- quantised DNN/CNN/RNN on FPGA/embedded systems per applicazioni online
- self-supervised (non-supervised) learning
- auto-ML
- ...



# STRUMENTI SVILUPPO

- formazione studenti:
  - dal 2020 corso di laurea in fisica / dal 2012 corso specialistico dottorato
- computing power e soprattutto stack software aggiornati e mantenuti



**16 TESLA V100 32 GB  
COMPLETAMENTE  
CONNESSE**

Memoria totale a elevata larghezza di banda da 0,5 TB per modelli di deep learning più complessi

**12 NVSWITCH**

Larghezza di banda bi-sezione da 2,4TB/s

**MEMORIA DI SISTEMA DA 1,5 TB**

Più memoria di sistema per gestire carichi di lavoro di deep learning di grandi dimensioni



teorico del flavor



fisico teorico o sperimentale  
deep learning enhanced ...

fisico teorico

