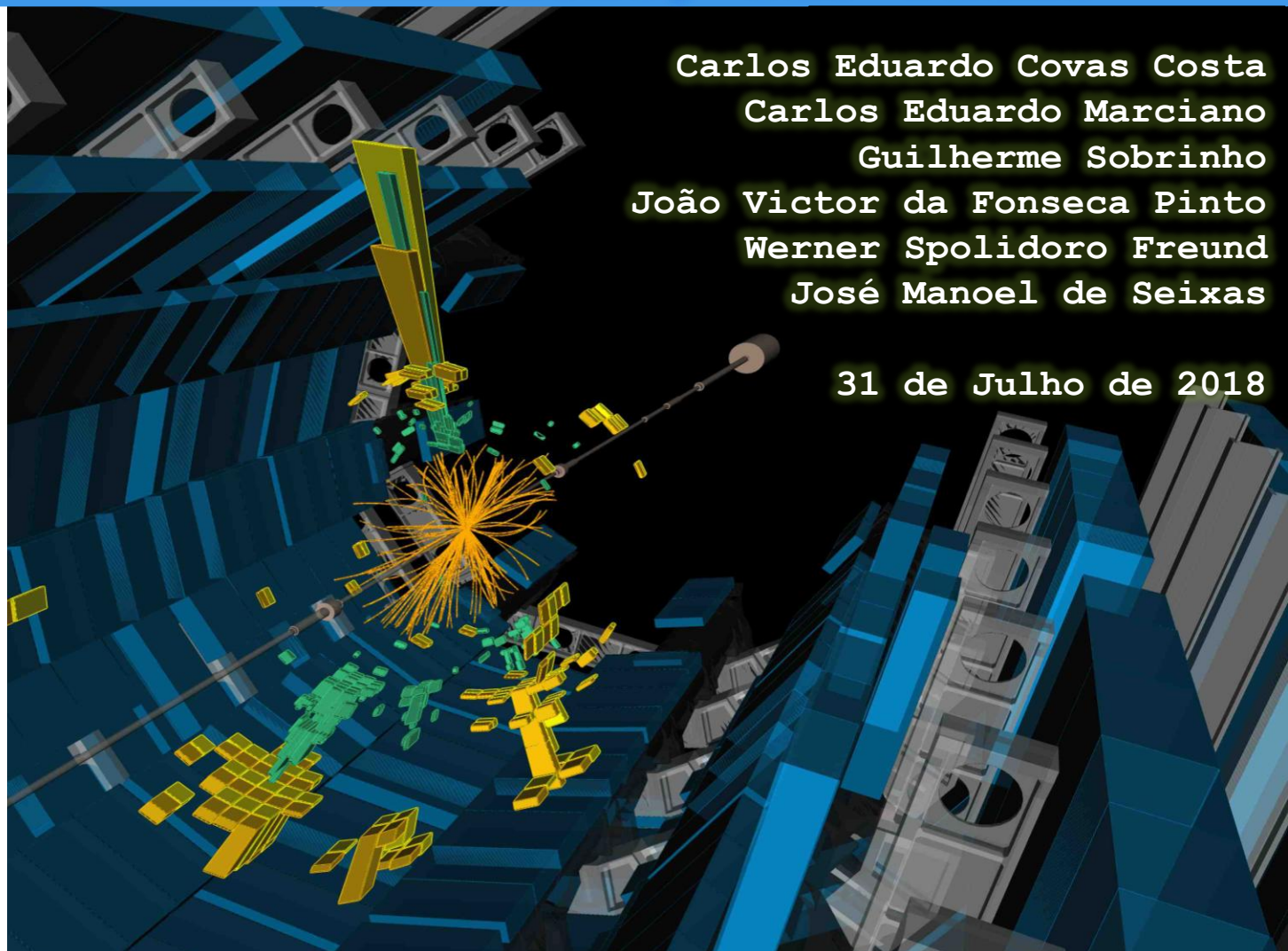
Filtragem Offline de Elétrons Baseada em um Ensemble de Redes Neurais Especialistas

RENAFAE 2018, São Paulo



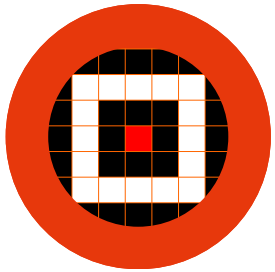
Carlos Eduardo Covas Costa
Carlos Eduardo Marciano
Guilherme Sobrinho
João Victor da Fonseca Pinto
Werner Spolidoro Freund
José Manoel de Seixas

31 de Julho de 2018

COPPE/Poli/UFRJ
Laboratório de Processamento de Sinais
Inteligência Computacional. Inovação

Agenda

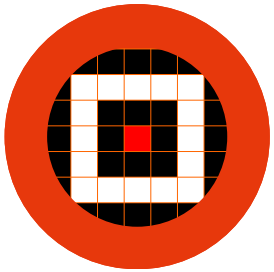
- Introdução
- Grandezas Físicas
- NeuralRinger
- Intervenção no Trigger
- Desenvolvimento no Offline
- Fusão de informação
- Resultados Offline
- Abordagens de pré-processamento
- Redução da massa de dados para treinamento
- Conclusões



Introdução

O ATLAS adotou o algoritmo NeuralRinger, que possibilitou aliviar a demanda por processamento durante a filtragem Online a partir de 2017.

- Nova descrição de calorimetria
  - Utilização de anéis concêntrios mantendo informação lateral e longitudinal.
- Ensemble de classificadores para tomada de decisão.
  - Utilização de ML no trigger para tomada de decisão principal.
- Não tinha contrapartida no Offline
  - Inversão da dinâmica do desenvolvimento.



Grandezas Físicas Discriminantes

7 Grandezas Físicas

Variables and Position

	Strips	2nd	Had.
Ratios	f_1, f_{side}	R_η*, R_ϕ	$R_{Had.}$*
Widths	$w_{s,3}$, $w_{s,tot}$	$w_{\eta,2}$*	-
Shapes	ΔE, E_{ratio}		

* Used in PhotonLoose.

Energy Ratios

$$R_\eta = \frac{E^{S2}_{3\times7}}{E^{S2}_{7\times7}} \qquad R_\phi = \frac{E^{S2}_{3\times3}}{E^{S2}_{3\times7}}$$

$$R_{Had} = \frac{E_T^{Had}}{E_T}$$

$$f_1 = \frac{E_{S1}}{E_{Tot.}}$$

$$f_{side} = \frac{E^{S1}_7 - E^{S1}_3}{E^{S1}_3}$$

Shower Shapes

$$E_{ratio} = \frac{E^{S1}_{max,1} - E^{S1}_{max,2}}{E^{S1}_{max,1} + E^{S1}_{max,2}}$$

$$\Delta E = E^{S1}_{max,2} - E^{S1}_{min}$$

Widths

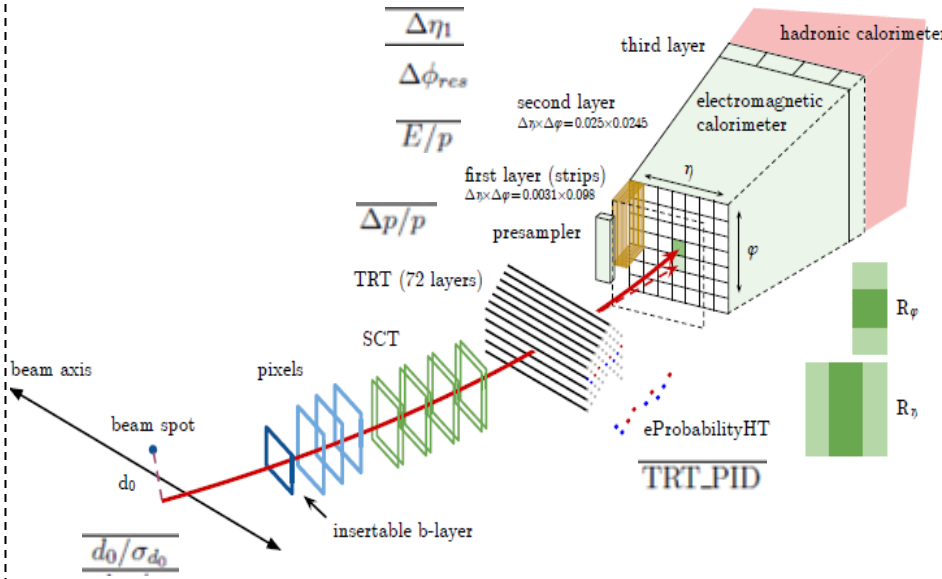
$$w_{\eta,2} = \sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i}\right)^2}$$

Width in a 3×5 ($\Delta\eta\times\Delta\varphi$) region of cells in the second layer.

$$w_s = \sqrt{\frac{\sum E_i (i - i_{max})^2}{\sum E_i}}$$

ws3 = w_s uses 3 strips in η; wstot is defined similarly, but uses 20 strips.

6 Grandezas Físicas



Sistema de Calorimetria (190 mil canais eletrônicos): absorção total da energia de elétrons, fótons e jatos hadrônicos.

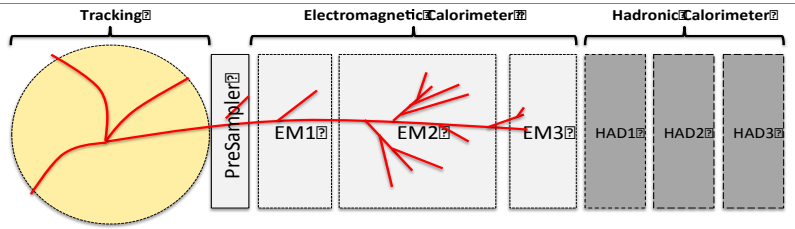
Dectetor de Interno (100 milhões canais): obtenção para partículas carregadas da trajetória, momento linear e outros.

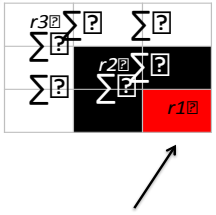
NeuralRinger

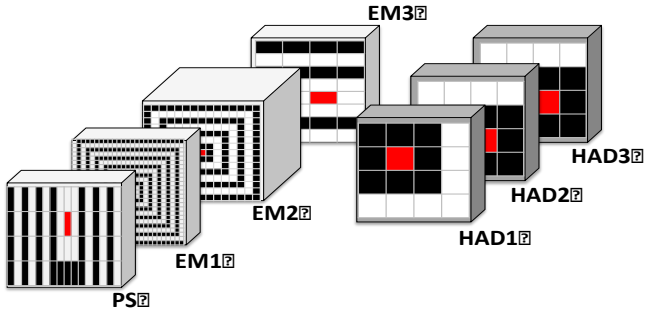
Alta dimensiona-lidade

- Proposta de longa data pela COPPE (1996);
- Explora a geometria cônica do chuveiro para descrever os padrões contidos na informação lateral e longitudinal, enquanto se obtém redução de dimensionalidade
- Generalização para boa parte das grandezas físicas.

100 Anéis de Energia



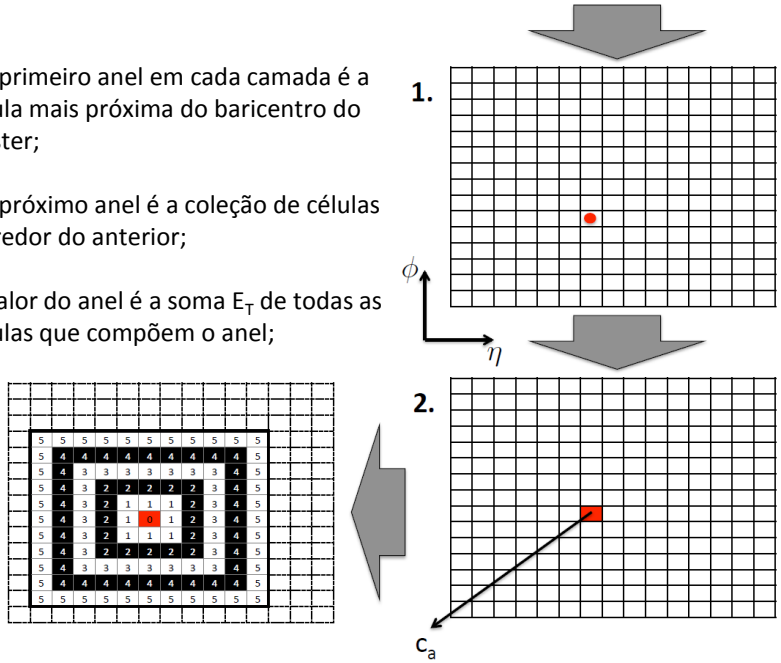




• O primeiro anel em cada camada é a célula mais próxima do baricentro do cluster;

• O próximo anel é a coleção de células ao redor do anterior;

• Valor do anel é a soma E_T de todas as células que compõem o anel;

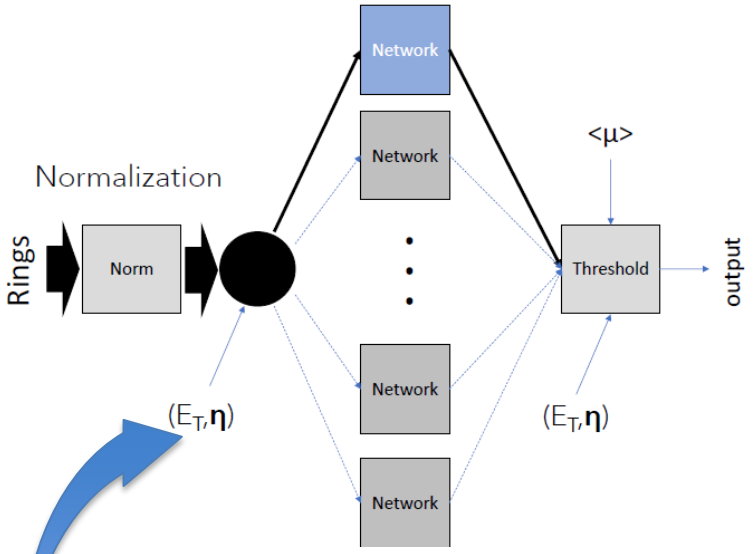


Número de anéis por camada						
PS	EM1	EM2	EM3	HAD1	HAD2	HAD3
8	64	8	8	4	4	4

Ringer Ensemble

• Da mesma forma que as grandezas de calorimetría, a representação anelada também está sujeita a distorções de acordo com a posição de interação e energia.

• Para lidar com estas distorções, O teste de hipótese é realizado através de um ensemble de classificadores compostos por MLPs especialistas em regiões do plano Et e η para discriminação elétron-hadron.



Inovação:
Introdução de técnica de ML para decisão no Trigger.

η \ E_T	0 : 15->20 [GeV]	1 : 20->30 [GeV]	2 : 30->40 [GeV]	3 : 40->60000 [GeV]
8 : 2.37->2.47	s: 28011 b: 190689	s: 127422 b: 108744	s: 245241 b: 25074	s: 321795 b: 16236
7 : 2.01->2.37	s: 112770 b: 753840	s: 536946 b: 429288	s: 1103685 b: 97104	s: 1562916 b: 62166
6 : 1.81->2.01	s: 64092 b: 415152	s: 319005 b: 235359	s: 694842 b: 54213	s: 1062351 b: 34851
5 : 1.52->1.81	s: 93663 b: 759504	s: 486054 b: 419529	s: 1129038 b: 92529	s: 1831269 b: 57987
4 : 1.37->1.52	s: 43962 b: 482994	s: 249855 b: 265143	s: 628254 b: 61113	s: 1055067 b: 37140
3 : 1.15->1.37	s: 69927 b: 619308	s: 409842 b: 314826	: 1064988 b: 69873	s: 1729662 b: 45978
2 : 0.80->1.15	s: 112686 b: 871245	s: 689847 b: 487542	s: 1978338 b: 113178	s: 3018663 b: 73500
1 : 0.60->0.80	s: 70551 b: 491217	s: 405981 b: 279036	s: 1243887 b: 65640	s: 1854609 b: 43359
0 : 0.00->0.60	s: 231150 b: 1422279	s: 1494546 b: 805794	s: 3735015 b: 191943	s: 5849436 b: 126174

Intervenção no Trigger

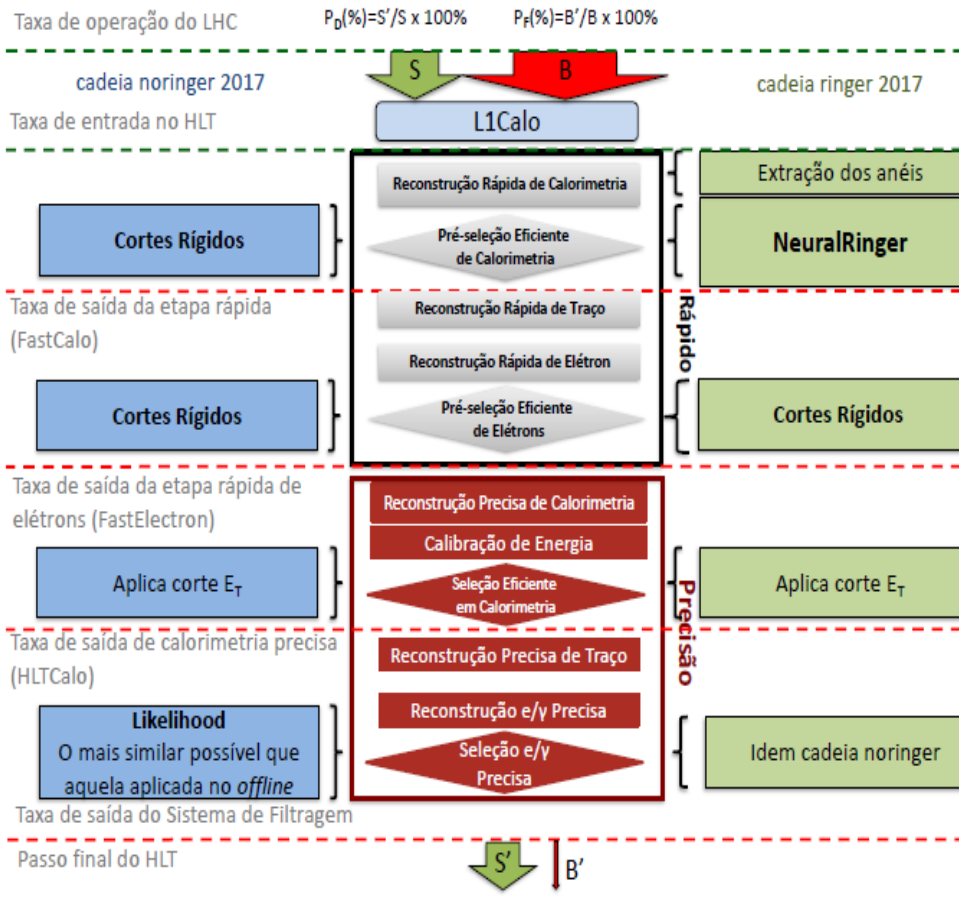
Problemas:

- Estudos em 2016 e início de 2017 mostraram que o consumo necessário para processamento nas condições de empilhamento não era sustentável.
- Reposição de parte da fazenda de processamento do HLT não foi realizada.
- Problema no setor 16L2 do LHC exigiu um novo esquema de preenchimento do LHC (8b4e).

Operação conforme previsões:

- Redução de 1/4 da demanda de processamento;
- No caso da cadeia principal: 200 ms/evento -> 100 ms/evento;
- Eficiência da cadeia praticamente inalterada;
- Ligeira tendência a coletar melhores elétrons conforme referência *offline*.

- Desenvolvido metologia de avaliação do impacto da intervenção no trigger:
  - A análise permite afirmar estatisticamente que há impacto e que o mesmo é positivo, resultando em um melhor perfil para o modelo da referência atuando no *offline*.

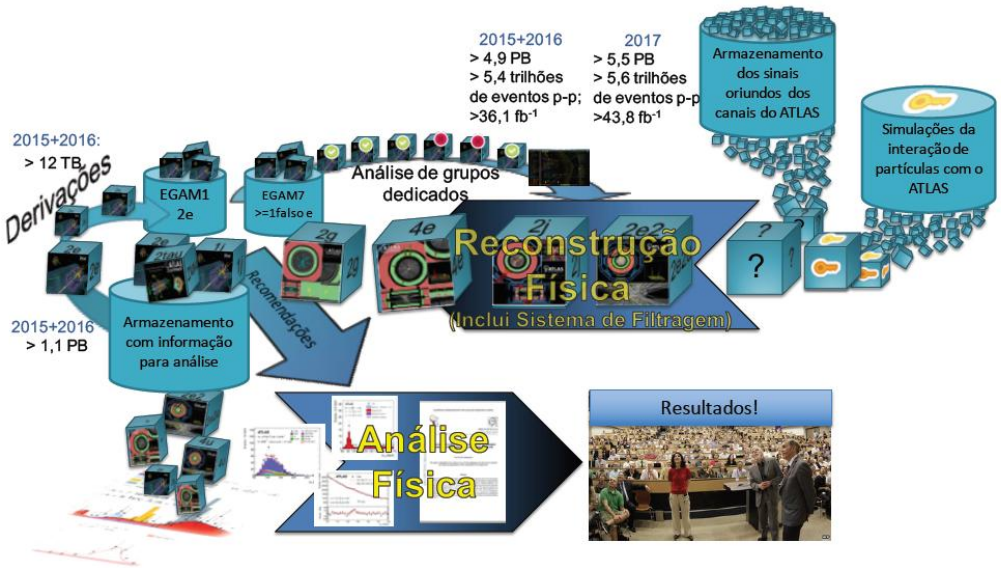


Otimização em eficiência no FastCalo devido à introdução do NeuralRinger:

- 2017: 159 <ms/evento>;
- Ringer: 121 <ms/evento>.

-24%!

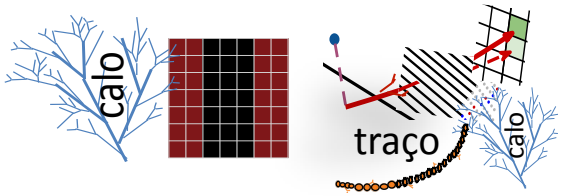
Desenvolvimento no Offline

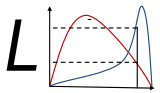


- Resultados no trigger mostram uma melhoria na qualidade das amostras no offline, em relação à técnica de cortes rígidos nas grandezas físicas, apesar do ensemble não ser alimentado pelas grandezas físicas.
  - Motiva aplicação em Offline
- Inversão da dinâmica do desenvolvimento.
- Offline permite estudo mais detalhado das nuances da técnica:
  - Estudo no trigger é realizado em referência ao *offline*.
  - Utilização de variáveis de traço para oferecer maior potencial discriminativo.
- Emprego de técnicas no trigger baseadas no *offline* possibilita maior segurança quanto à natureza de seu comportamento.
- Uso de técnica baseada em Aprendizado de Máquinas, não-usual no contexto de identificação de elétrons.
- Possibilidade de estudo de modelos mais complexos e busca de um espaço latente mais discriminante a partir da representação dos anéis.

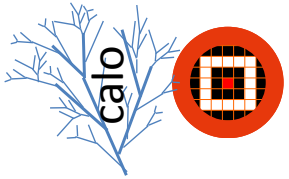
Fusão de Informação

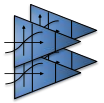
- Operação *online* limitada à informação de calorimetria;
- Necessário integrar informação de traço para permitir o acesso à toda informação discriminante.





Verossimilhança

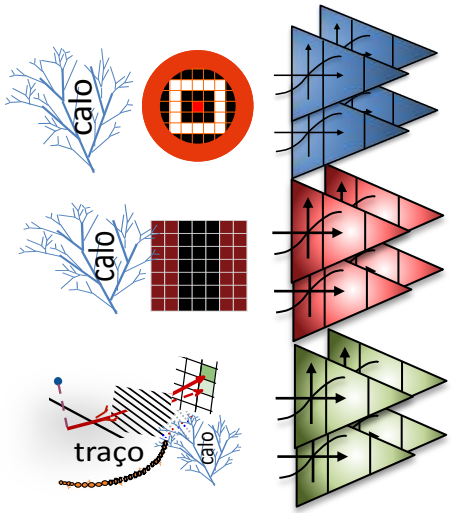




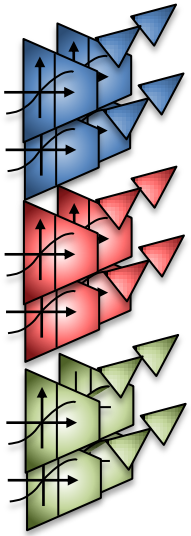
NeuralRinger

Redes Neurais Especialistas

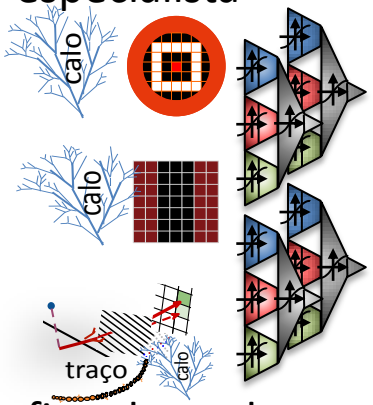
1. Treino de rede especialistas compondo o ensemble em separado



2. Excluir a camada de saída com os pesos para formação do espaço de saída



3. Treino da rede de fusão a partir dos padrões obtidos no treino especialista



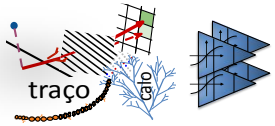
Ajuste fino das redes especialistas ganho de eficiência

Método Offline

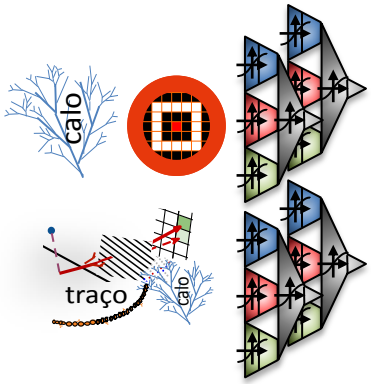
- Método desenvolvido a partir implementado no online, aproveitando a informação de traço disponível do ID.
- Desenvolvimento em dados de simulação contendo condições similares àquelas encontradas em colisões de 2016.
- Avaliação apenas do critério veryloose:
  - Técnicas novas devem ser mais necessárias em casos mais extremos.
  - Física Nova
- Emprego das mesmas fronteiras para ajuste de parâmetros que a referência (LLH)- 36 espaços.
- Avaliação das redes especialistas *per se* e da rede de fusão dos anéis com as grandezas de traço.
- Algumas grandezas físicas de calorimetria não são generalizáveis a partir dos anéis:
  - Avaliação da introdução dessa informação nas redes especialistas compondo o ensemble.

NN(rings)

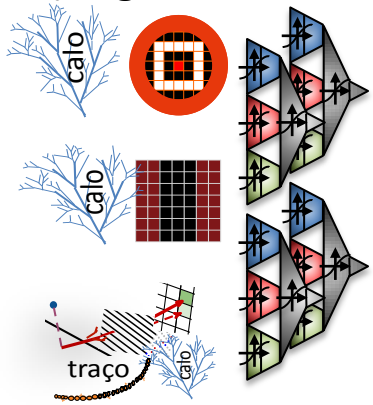
NN(track)



ExNN(rings + track)

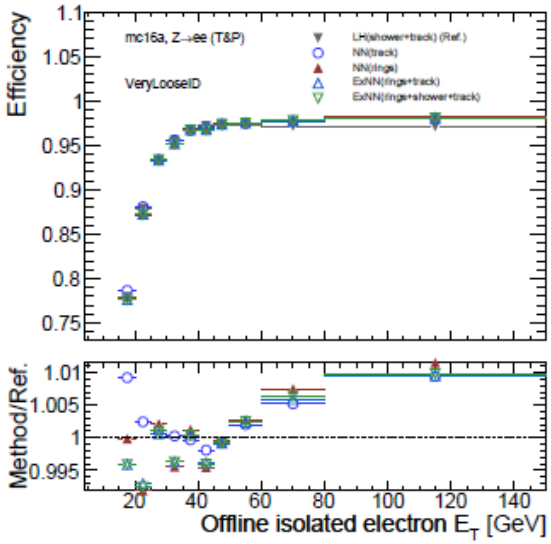


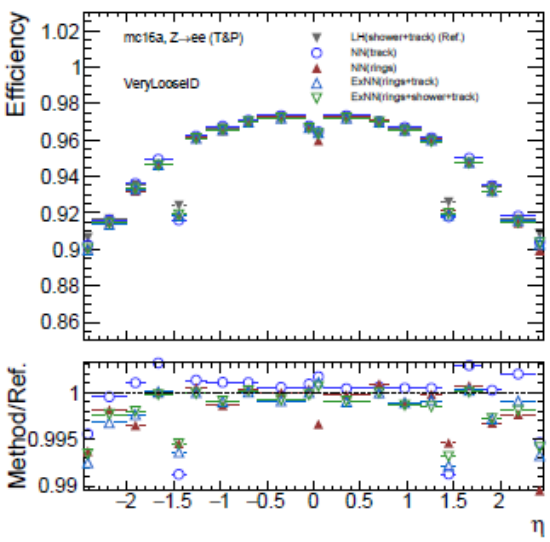
ExNN(rings + shower + track)



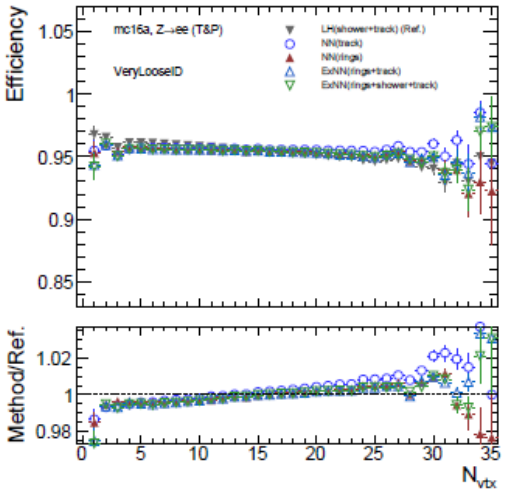
Resultados: Eficiência e Falso Alarme

Eficiência em T&P

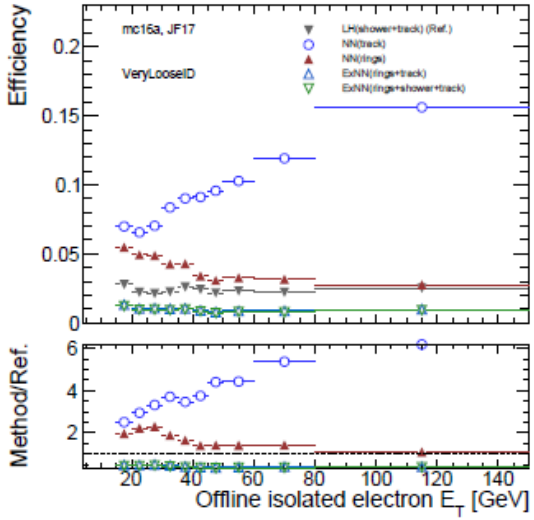


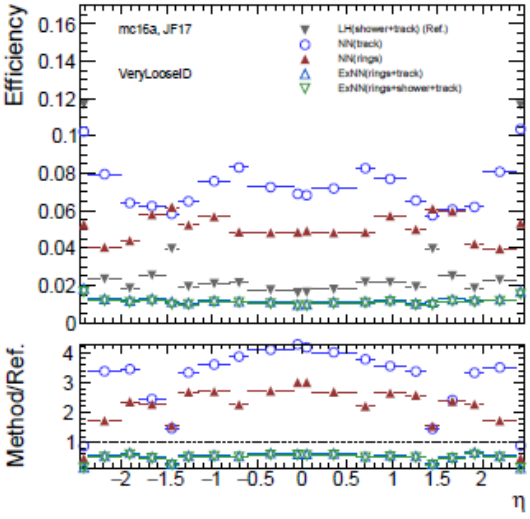


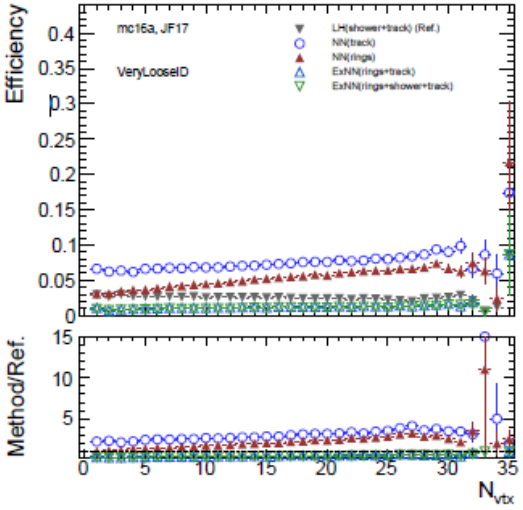
Efeito de empilhamento



Falso Alarme







Resultados Offline

Ponto de operação: mesma eficiência que referência;

Method	Type	$P_D[\%]$	$SP[\%]$	$F_a[\%]$
$LH(shower+track)$	Reference	96.82	97.14	2.54
$NN(rings)$	CrossValidation	96.72±0.00	95.80±0.00	5.12±0.00
$NN(track)$	CrossValidation	96.80±0.00	94.77±0.00	7.24±0.00
$ExNN(rings+track)$	CrossValidation	96.75±0.00	97.76±0.00	1.22±0.00
$ExNN(rings+shower+track)$	CrossValidation	96.75±0.00	97.81±0.00	1.13±0.00

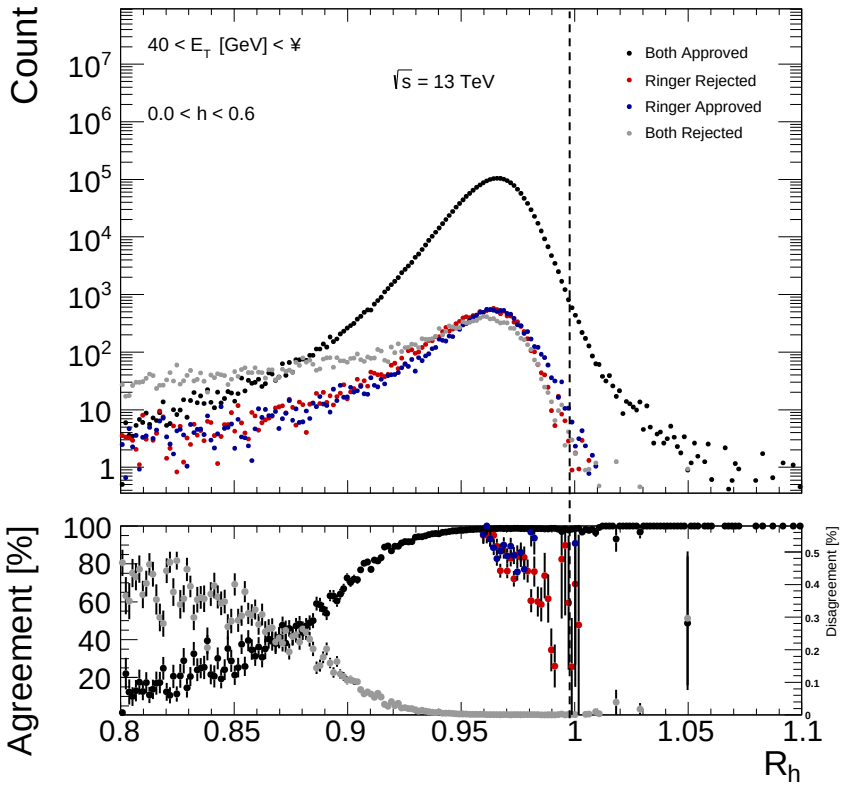
- Resultado integrado para todas as regiões atuantes dos ensembles;
- Erros de validação cruzada são desprezíveis;

- Redução do falso alarme em relação à referência por um fator de 2x;
- O ensemble com fusão das grandezas físicas de calorimetria apresenta uma redução de 0,1% no falso alarme em relação ao ensemble sem essa informação.

Análise de Quadrante

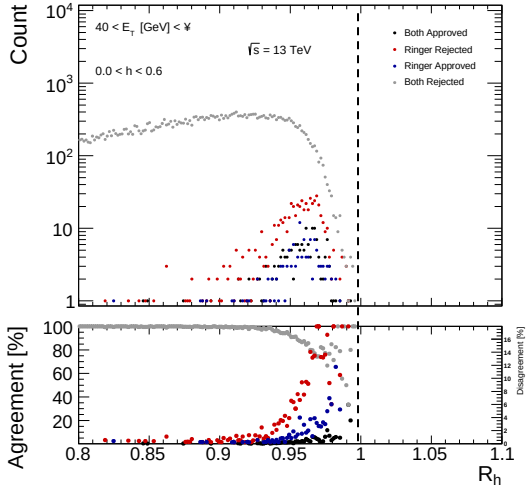
- Análise de quadrante em sinal mostra ligeira tendência a coletar amostras com melhor R_n mesmo para o caso do ensemble sem ser alimentado com essa informação;

R_η: próximo à unidade indica menor vazamento lateral (elétron)



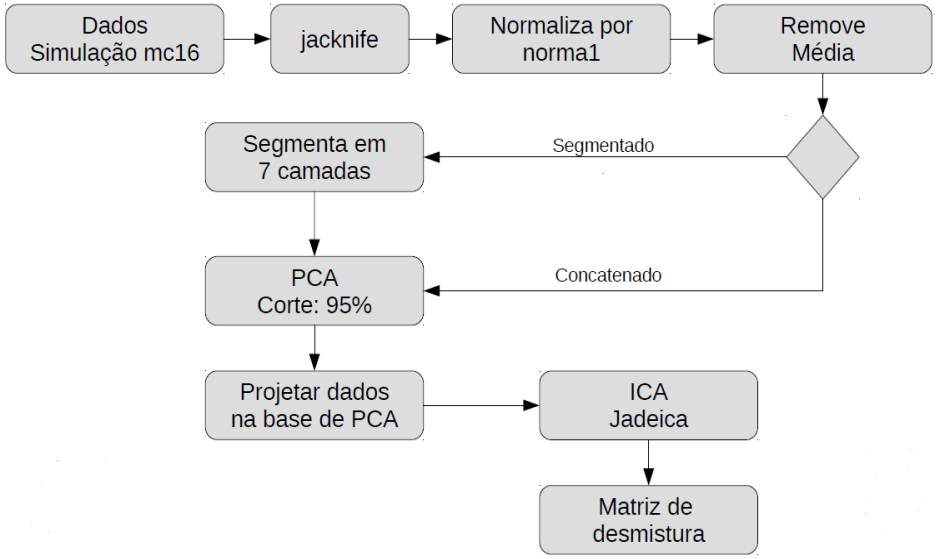
- Análise de quadrante em falsos elétrons mostra tendência esperada: maior falso alarme para região com perfil próximo ao de elétrons.

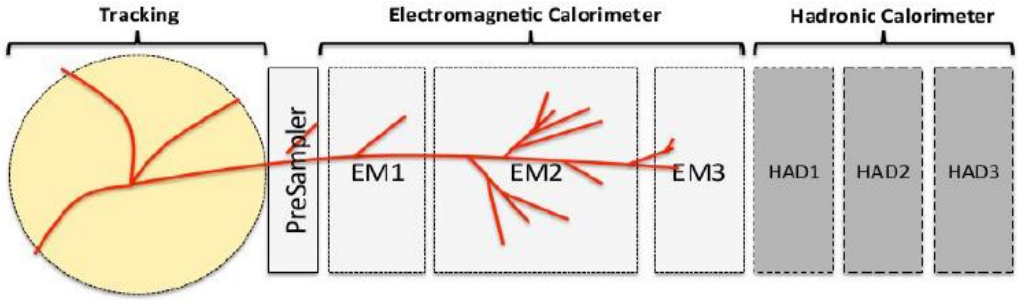
- Não é possível afirmar se há diferença no perfil dos falsos elétrons (ideal avaliação para ponto de operação no mesmo falso alarme).



Pré-processamento: ICA

- ICA procura encontrar uma nova base de representação buscando obter independência estatística entre as variáveis nova representação.
  - Fontes mantidas após o pré-processamento por anelamento da informação.
- Abordagem segmentada e concatenada.
  - ICs de cada camada => inf. lateral
  - ICs concatenada => inf. longitudinal
- Acredita-se que o número de ICs seja menor que a dimensão dos anéis.
- PCA acessa estatística de segunda ordem:
  - Remove correlação linear entre os padrões
- Base de projeção:
  - Autovetores da matriz de covariância.
  - Autovalores indicam variância de cada componente.
  - Maximizam a variância das projeções
  - São ortogonais na direção de maior variância.





Pré-processamento: Stacked Autoencoders

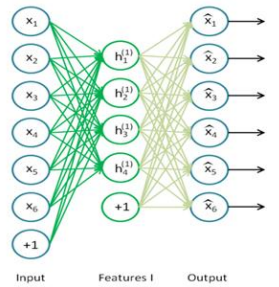
• Autoencoders são redes neurais que aprendem codificações eficientes de maneira não supervisionada.

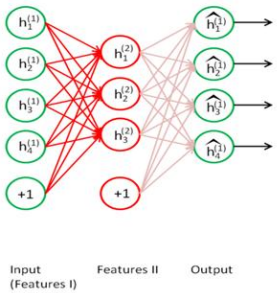
• Permite compressão não-linear da informação anelada de calorimetria.

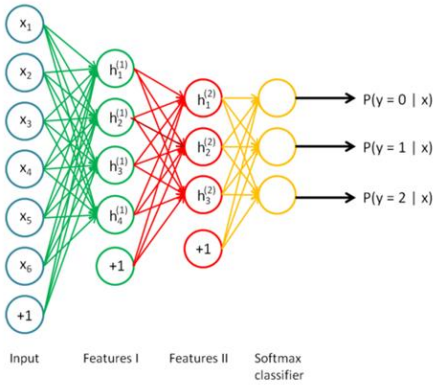
  • Obter representação mais discriminante do desenvolvimento do chuveiro.

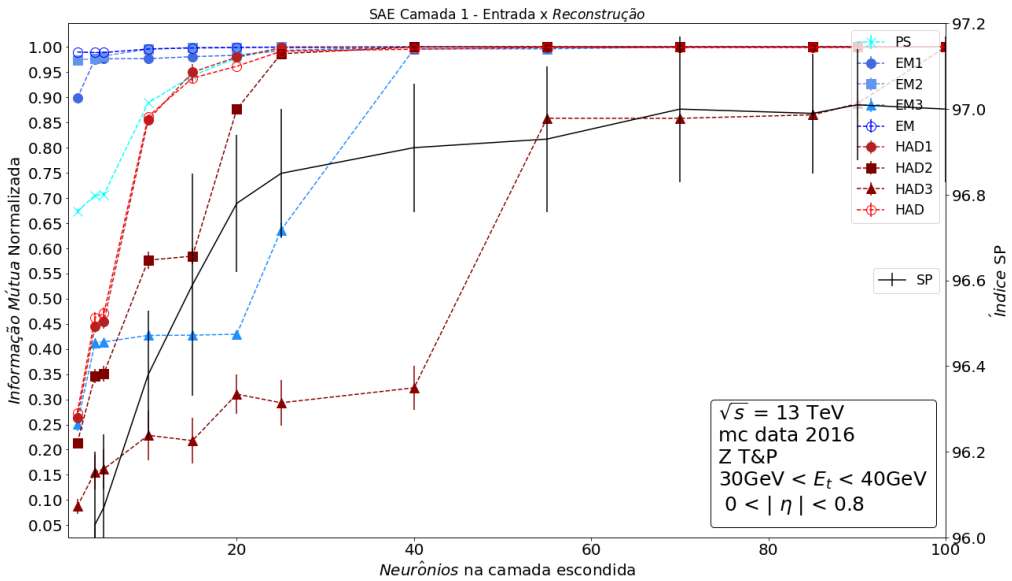
- Evitar função identidade -> Restrições:
  - Undercomplete
  - Overcomplete
  - Sparse
  - Denoising
  - Contractive

- Treinamento pode ser feito em 2 etapas:
  - primeiro de forma não-supervisionada
  - depois de forma supervisionada (fine tuning)









Pré-processamento: Ringer RP

• Estudos realizados nos SPACAL para avaliar discriminação elétron-píon empregaram o valor de Rp com valores unitários de α e β em dados de feixes experimentais.

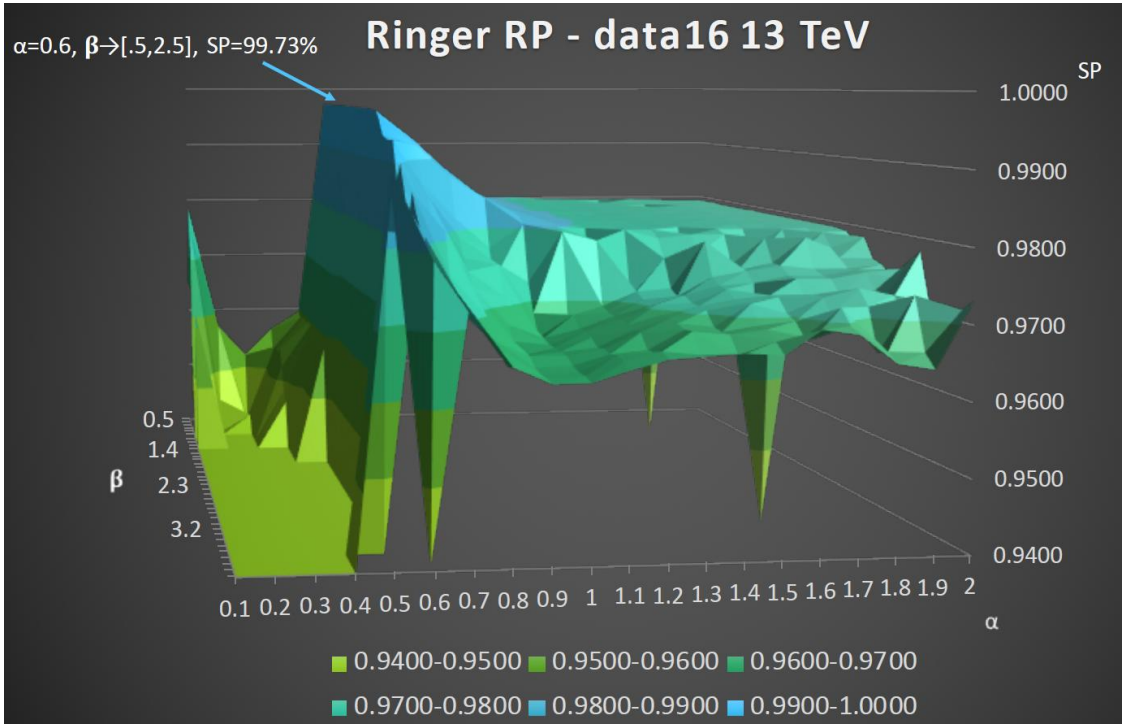
• Ringer-RP consiste em uma forma de explorar o espaço latente utilizando conhecimento especialista.

• Realiza-se uma aproximação de R_p ao aplicar o mapeamento não-linear no espaço de fase descrito pela energia dos anéis e suas respectivas distâncias em relação ao centro.

- Os parâmetros α e β são otimizados através de busca em grade,

$$R_p = \frac{\sum_{i=1}^{n} E_i^{\alpha} r_i^{\beta}}{\sum_{i=1}^{n} E_i^{\alpha}}$$

- Onde r_i é a distância euclidiana do centro da i-ésima célula ao baricentro de energia do chuveiro.



Reduzindo a massa de dados do Treinamento

• Grande massa de dados (> 1 milhão) leva longos ciclos de treinamento (> 24hs) para testar melhorias.

• Estatística disponível deve crescer de forma mais rápida que os recursos destinados ao processamento.

• Modelos se tornam mais complexos ao mesmo tempo que mais observações estão disponíveis.

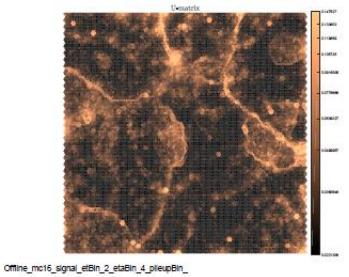
Como selecionar com quais registros treinar?

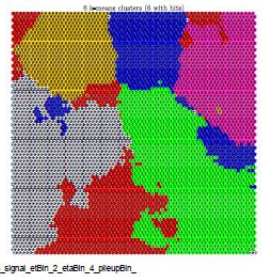
• Em desenvolvimento: técnicas que focam em manter a distribuição espacial dos dados e dar prioridade à fronteira entre classes.

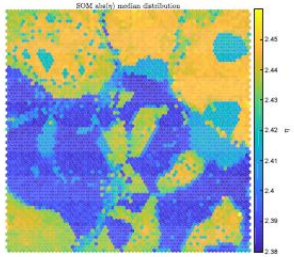
  • Reduzir informação redundante para obter menor demanda computacional.

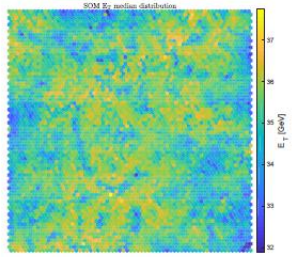
  • Dividir o espaço em grupos que irão compor o subconjunto para participar do processo de treinamento.

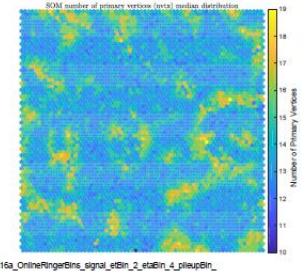
Mapas auto-organizáveis

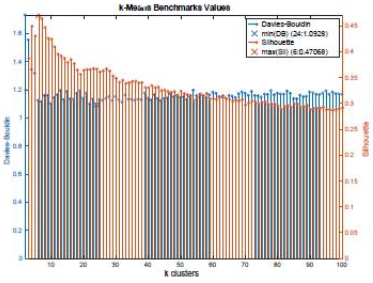










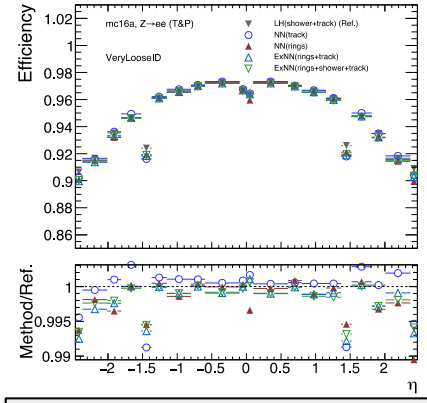


Conclusões

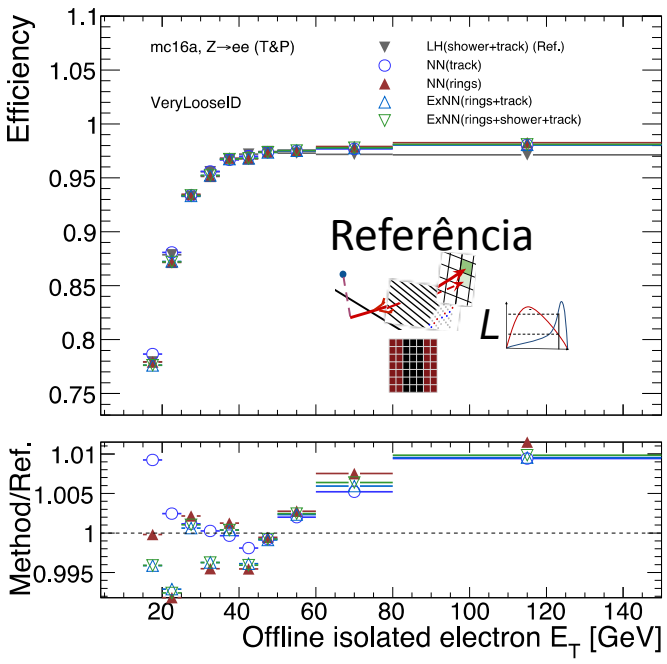
- Infraestrutura:
  - Intervenção realizada na reconstrução offline que se encontra atualmente em atuação após obter impacto desprezível no processamento.
  - Disponíveis 10 PB de dados de colisão (+2018 quando disponível) e respectivas simulações correspondentes processados que possibilitam o desenvolvimento completo de técnicas para identificação de elétrons.
- Eficiência:
  - Avaliação em simulação mostra bom potencial da extensão do NeuralRinger para offline.
  - Para mesma eficiência, redução no falso alarme (2,54->1,13%) e ligeira coleta de amostras de melhor qualidade para a física.
- Pré-Processamento:
  - Outras abordagens para encontrar um espaço latente mais discriminante a partir da representação dos anéis empregando:
    - pré-processamento estatística através de ICA.
    - Conhecimento especialista através de mapeamento não-linear (Ringer-Rp).
    - Compactação não-linear cega via SAE.
- Big data
  - Desenvolvida estratégia para reduzir o processamento necessário no desenvolvimento dos modelos avaliados, enquanto mantendo as amostras relevantes para o ajuste.

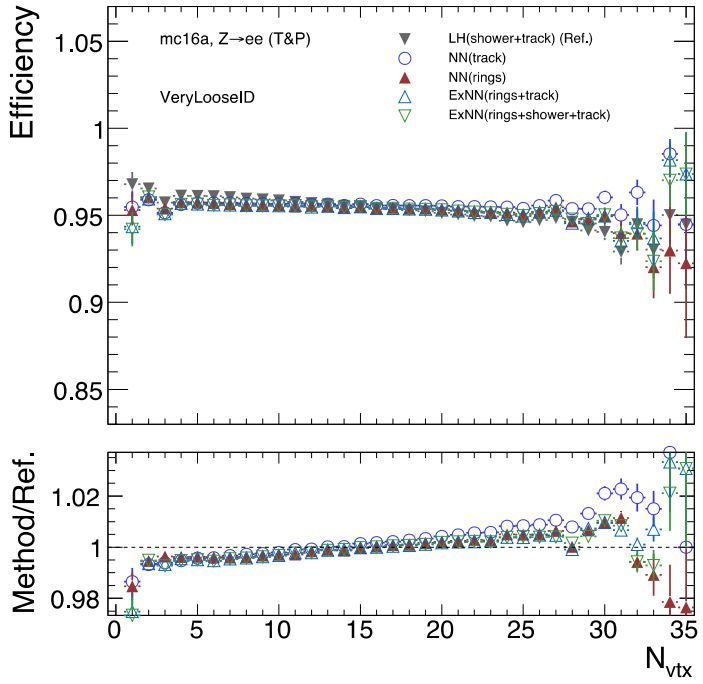
Backup

Resultados: Eficiência



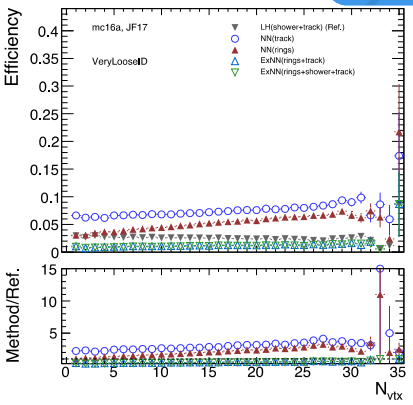
Ponto de operação: mesma eficiência que referência;

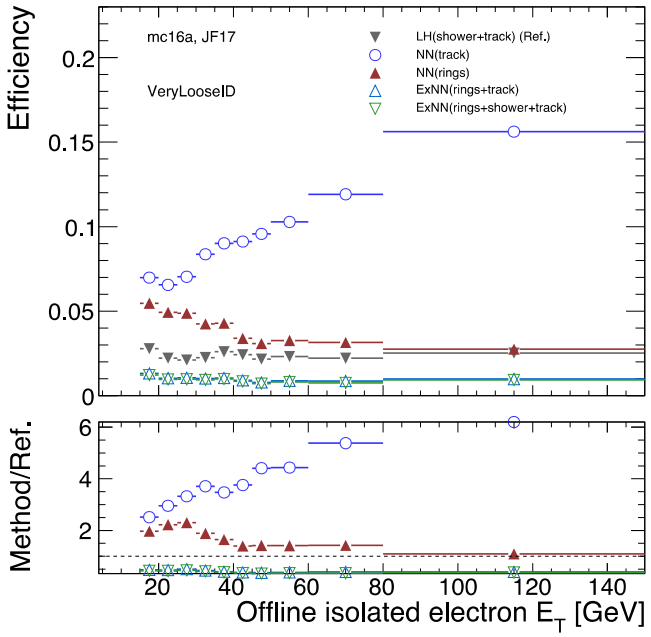


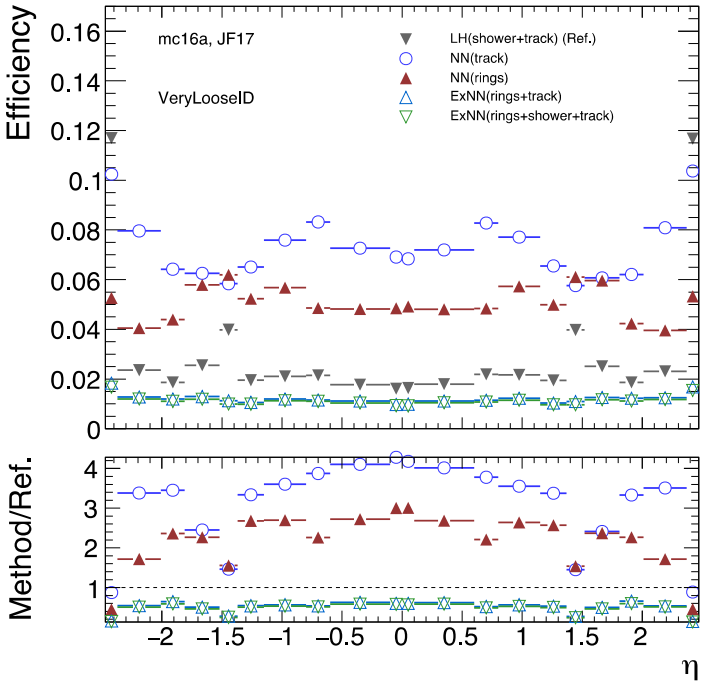


- Em todos os casos oscilações em relação à eficiência geralmente inferiores à 1%;
- Ligeira tendência dos métodos à obter menor dependência em função de empilhamento:
  - Referência ajustada em dados de colisão, que contém maior dependência.
- Maior eficiência em regiões superiores de energia:
  - Interessante para a busca de nova física.

Resultados: Falso alarme







Ponto de operação: mesma eficiência que referência;

- Ensemble especialistas apresentam alta influência em função do empilhamento;
- Referência mostra queda de falso alarme conforme empilhamento reforçando observação anterior;
- Ensembles com fusão apresentam falso alarme estável inferior para todas regiões de nxE_T;
- Crescimento de falso alarme para ensemble especialista em traço de acordo com o conhecimento especialista;
- Ganhos na região do crack (1,37<|n|<1,52) devem ser mais modestos que os apresentados.