

Combinando Redes Neurais Artificiais e Representação Esparsa de Dados para Reconstrução Online de Energia no Calorímetro Hadrônico do ATLAS

Tiago M. Quirino, Luciano M. A. Filho e José M. Seixas

Resumo— Exige-se o processamento de grandes quantidades de dados gerados pelos muitos canais de medição de energia do experimento ATLAS no acelerador de partículas LHC. Os canais de leitura fornecem um sinal alterado em relação ao evento físico, isso combinado ao aumento da taxa de eventos, planejado para o LHC, fazem surgir o efeito de empilhamento do sinal. Técnicas de deconvolução do canal tem sido pesquisadas, existindo interesse nas redes neurais artificiais (RNA) e representações esparsas do sinal. Logo, neste trabalho, propõe-se utilizar a unidade linear retificada (*reLu*) na RNA, combinando as técnicas de RNA e representação esparsa do sinal.

Palavras-Chave— Redes Neurais Artificiais, Calorimetria, Esparsidade.

Abstract— It requires the processing of large amounts of data generated by the many energy measurement channels of the ATLAS experiment on the LHC particle accelerator. The reading channels provide a changed signal in relation to the physical event, combined with the increase in the event rate, planned for the LHC, give rise to the signal pileup effect. Techniques of deconvolution of the channel have been researched, existing interest in artificial neural networks (ANN) and sparse representations of the signal. Therefore, in this work, it is proposed to use the rectified linear unit (ReLU) in ANN, combining ANN techniques and sparse representation of the signal.

Keywords— Artificial Neural Networks, Calorimetry, Sparsity.

I. INTRODUÇÃO

Experimentos inovadores de pesquisa científica exigem sistemas de instrumentação sofisticados, pois a falta de dados comparativos impõe a necessidade de garantir confiabilidade dos dados medidos e assim permitir a correta interpretação das descobertas. O desenvolvimento da pesquisa em Física de Altas Energias se enquadra no contexto de vanguarda científica, especificamente a aceleração e colisão de feixes de partículas, no intuito de explicar melhor a constituição da matéria e suas interações [1].

Calorímetros são alguns dos equipamentos instalados nas regiões de colisão dos feixes de partículas acelerados, para que a energia dos subprodutos das colisões seja mensurada e as partículas sejam caracterizadas. Estas depositam toda sua energia progressivamente na forma de cascata de partículas

Tiago Motta Quirino, Universidade Federal de Juiz de Fora, Juiz de Fora - MG, e-mail: tiago.m.quirino@gmail.com; Luciano Manhães de Andrade Filho, Universidade Federal de Juiz de Fora, Juiz de Fora-MG, e-mail: luciano.ma.filho@gmail.com, José Manoel de Seixas, Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, e-mail: seixas@lps.ufrj.br.

até o limiar de detecção [2] em períodos da ordem de nanosegundos. Demanda-se, nos calorímetros, um sistema de instrumentação, tanto em *hardware* como em *software* complexo, pelos requisitos de tempo e acurácia [3]. Principalmente considerando altas taxas de eventos, em que os sinais medidos podem ser sobrepostos se o tempo de resposta de detecção não for suficientemente pequeno.

No cenário de aceleração de partículas, o LHC (*Large Hadron Collider*) é atualmente o acelerador em operação com as maiores dimensões e capaz de fornecer mais energia, colidindo feixes de prótons em uma taxa de 40 milhões por segundo.

Existem quatro experimentos instalados para detecção de partículas: ALICE, LHCb, CMS e ATLAS [4]. O ATLAS (*A Toroidal LHC Apparatus*) é um dos experimentos de detecção de partículas constituído por subsistemas, no entorno do ponto de colisão, como indicados na figura 1: detector de trajetória, calorímetro eletromagnético, calorímetro hadrônico e o espectrômetro de múons [5].

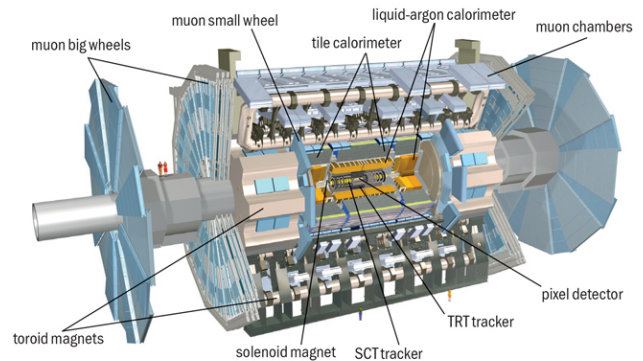


Fig. 1. Subsistemas de detecção do ATLAS em camadas. Extraído de [5]

Para os próximos anos, são previstas atualizações graduais no LHC elevando a densidade de feixes de prótons, que no âmbito de física de altas energias é chamado de aumento de luminosidade do acelerador [6]. Consequentemente, a taxa de interações de partículas e a quantidade de partículas elementares detectadas aumenta, de tal maneira a permitir maior probabilidade de descobertas de partículas elementares.

Com o aumento de luminosidade os métodos de processamento de sinais atuais se tornam ineficientes, em espe-

cial considerando os calorímetros do ATLAS. Recentemente, métodos que consideram as características não-lineares do ruído de fundo [7] e métodos que exploram a estrutura esparsa dos sinais a serem reconstruídos [8] nos calorímetros, vem demonstrando um melhor desempenho neste ambiente de alta luminosidade do que os métodos tradicionalmente usados em etapas anteriores.

O calorímetro hadrônico do ATLAS - Tilecal (do inglês, *Tile calorimeter*) -, é composto por material pesado (aço) que interage com o chuveiro ou cascata de partículas resultantes das colisões e por telhas cintilantes, nas quais as partículas depositam parte de sua energia que é convertida em luz. Totalizam-se 10.000 canais de medição [9], que são capazes de detectar a luz produzida pelas telhas, o que permite boa localização do chuveiro de partículas. A luz detectada é convertida em sinal elétrico e enfim se realiza o condicionamento e digitalização do sinal que deve possuir amplitude proporcional à energia medida [10]. Recuperar a amplitude de energia depositada na célula é o interesse nesta etapa, através da estimação a partir do sinal proporcional adquirido.

A janela de medição do Tilecal é de 150ns que representa todo sinal do pulso [11], mas por outro lado o evento físico possui um período de 25ns, no caso da taxa de eventos ser baixa existe intervalos entre as colisões. Nas atualizações futuras estima-se que com o aumento de energia proposto para o LHC a taxa de interações por colisão no calorímetro aumente de 60 para 200. Por outro lado, não se propõe aumentar a velocidade de resposta do medidor [12], logo, deposições sucessivas de energia em altas taxas acarretam maior possibilidade de sobreposição de sinais subsequentes na mesma janela de medição, ou seja, aumento da ocupação de eventos na janela, que no contexto da calorimetria é denominado ocupância, isso produz o efeito de empilhamento da medição (*pileup*) representado na figura 2.

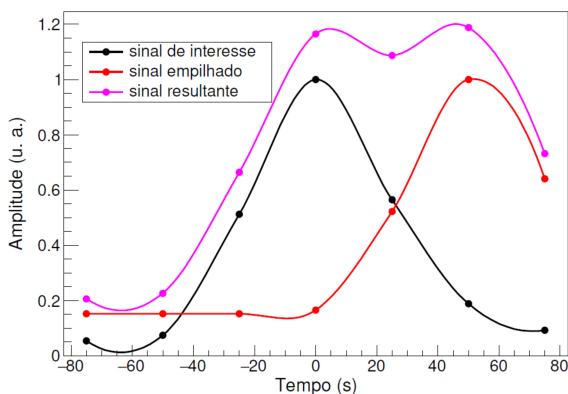


Fig. 2. Efeito do empilhamento. Extraído de [3]

Introduz-se assim, um desafio ao processamento dos sinais obtidos dos canais, pois os métodos utilizados atualmente não são capazes de tratar o efeito de *pileup* [3].

A interpretação dos canais de detecção do calorímetro como canais de comunicação, permitiu modelar a estimação da energia como o resultado da deconvolução do canal com o sinal de saída do medidor, ou filtragem inversa. Em trabalhos anteriores

foram propostas duas considerações para reconstrução do sinal nos canais de medição, usando o princípio de filtragem inversa: a representação mais esparsa quanto possível do sinal detectado [8]; e a natureza não-linear do canal de leitura do Tilecal, pelo ruído de fase proveniente do tempo de voo das partículas [7].

Portanto há interesse na aplicação de um método que adote a esparsidade e a não-linearidade em conjunto, características não consideradas no método atual e ainda não combinadas em trabalhos anteriores para tal aplicação.

II. MÉTODO PROPOSTO

Considerando o modelo do sinal medido pelo Tilecal sendo o resultado da convolução de amostras de energia e o pulso característico do canal, deseja-se reconhecer os padrões do efeito do canal no processo de convolução para reconstrução do sinal de energia.

As amplitudes de energia do evento ($a[n]$) e os valores medidos do Tilecal ($x[n]$) são obtidos em pares e representados por vetores de mesma dimensão $\{x, a\}$, pois são variáveis síncronas, em que n representa o índice de amostras do ADC.

Os algoritmos propostos, até o momento, para reconstrução do sinal de energia são treinados de forma offline, quando utilizados dados obtidos de simulações de Toy Monte Carlo que reproduzam as características estatísticas do ambiente físico, a saber, uma distribuição de probabilidade exponencial das amplitudes de energia das colisões medidas [13]. Após o treinamento, o algoritmo que apresente resultados melhores na simulação é implementado em FPGA para processamento online do sinal oriundo dos canais de medição do tilecal.

Neste trabalho, foram gerados conjuntos de dados simulados com diferentes ocupâncias variando de 10% a 90%, com interesse de representar os diferentes padrões de *pileup*. Para cada conjunto de dados ao ordenar os valores de amplitude, tanto dos eventos $a[n]$, quanto das medições $x[n]$, seguem o padrão de regressão representado no exemplo de 30% de ocupância da figura 3. Percebe-se que existe a característica não-linear para pequenos valores de amplitudes, uma região intermediária aproximadamente linear e uma região não-linear considerando amplitudes mais altas.

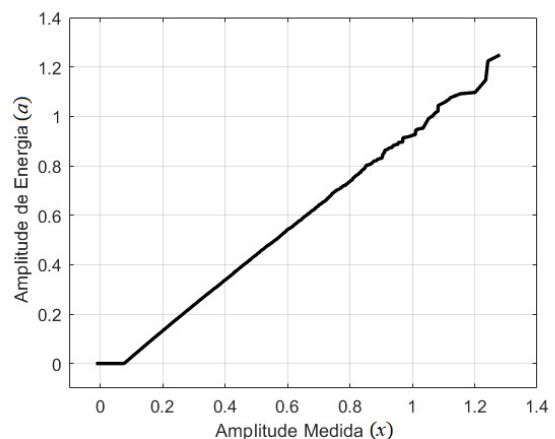


Fig. 3. Exemplo de mapeamento da entradas e saídas ordenadas para 30% de ocupância.

A. Reconstrução do Sinal pela Abordagem de Esparsidade

Em [8] utilizou-se a abordagem de deconvolução do canal de leitura do Tilecal ($h[n]$) por aproximação de um filtro FIR ao inverso da resposta impulsiva do canal. Além disso combinou-se a abordagem por esparsidade, que considera uma função não-linear para determinar uma solução que considere a maior esparsidade possível do sinal. O sistema resultante é um filtro FIR de ordem 12 ($m = 12$) na entrada em série com uma função denominada Shrinkage, como apresentado na figura 4. No intuito de fornecer mais informações do sinal ao filtro, há interesse no sinal central do conjunto de amostras armazenado no *buffer*, por isso estima-se a amostra com atraso $\frac{m}{2}$, considerando a quantidade de amostras par.

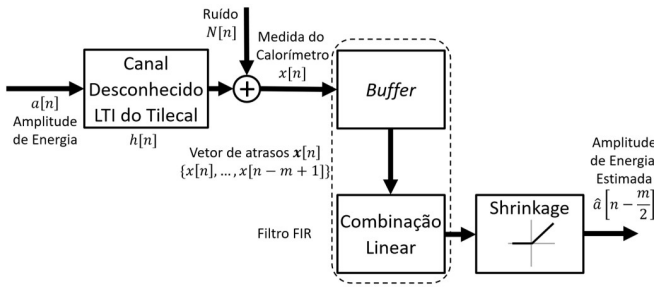


Fig. 4. Sistema FIR considerando a esparsidade do sinal.

B. Reconstrução do sinal por RNA

Considerando outra perspectiva, a modelagem dos canais de comunicação é realizada pela utilização das Redes Neurais Artificiais (RNAs) [14]. Devido à possibilidade das RNAs realizarem aproximação universal de funções com um erro arbitrariamente pequeno, pois separa padrões entre variáveis, desde que disponibilize-se para a parametrização das RNAs uma quantidade de dados acima da ordem de milhar. Pela capacidade de generalização, conjuntos de dados que não foram aprendidos pela rede ainda geram respostas com erro pequeno, por isso é possível que se aplique às RNAs um banco de dados incompleto, ambíguo e parcialmente incorreto [15].

A modelagem do canal de comunicação do Tilecal foi realizada em [7], utilizando a topologia *Multi Layer Perceptron* (MLP), com função de ativação tangente hiperbólica na camada oculta e uma combinação linear na camada de saída. Os pesos sinápticos interligando as camadas armazenam os padrões de entrada-saída e o aprendizado de padrão através do ajuste dos mesmos foi realizado pelo algoritmo de Levenberg-Marquardt. Utilizando como critérios de parada a checagem por validação cruzada, evitando o *overfitting*. Empiricamente foi decidida a estrutura final da rede com 10 entradas que são uma janela de atrasos do sinal do canal de leitura ($m = 10$) e uma camada escondida com 4 neurônios ($h = 4$), cujo sistema resultante é apresentado na figura 5. Mantém-se o argumento de estimar a amostra central para que a rede possua mais informações do sinal, ou seja, estimar a amostra de amplitude de energia $a[n - \frac{m}{2}]$.

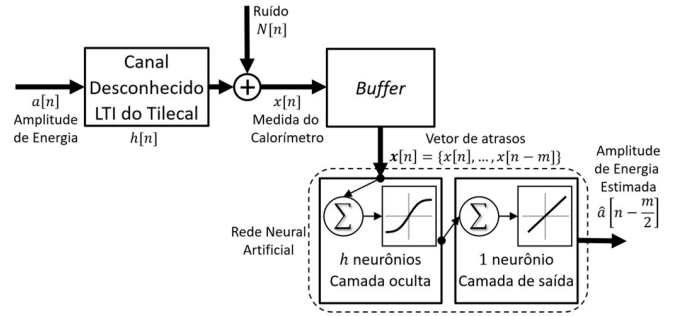


Fig. 5. Sistema com implementação da função ReLU na RNA.

C. Esparsidade em Redes Neurais Artificiais

Analisando em conjunto as figuras 4 e 5, percebe-se que é possível combinar as vantagens de cada método, a saber, a função não-linear Shrinkage na saída do filtro FIR da figura 4, que promove a esparsidade, com o processamento não-linear da camada oculta da rede neural da figura 5. É interessante, neste ponto, destacar que a função Shrinkage é idêntica, salvo a nomenclatura em cada contexto, a função de ativação linear retificada (ReLU, do inglês *Rectified Linear Unit*), que vem sendo cada vez mais usada em RNA [16].

A função ReLU é definida como a parte positiva do seu argumento, ou seja, linear em tal intervalo, e zero no intervalo negativo, o que atribui não-linearidade à função, conforme a equação (1) e como apresentado na figura 6. Um neurônio implementado com a função de ativação ReLU é considerado desativado na ocorrência de um valor menor que zero da combinação linear dos pesos e entradas respectivas, ou seja, o mesmo não influencia a saída. Portanto, alguns neurônios não são ativados quando determinados valores de entrada são introduzidos à RNA, tal condição, de ativação ou não de neurônios torna a RNA esparsa [17].

$$ReLU(u) = \max(0, u) = u^+ \quad (1)$$

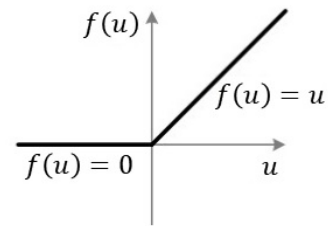


Fig. 6. Função de Ativação linear retificada.

Dentre as vantagens da função ReLU destacam-se: a facilidade de ajustamento dos pesos, ao não ativar todos os neurônios ao mesmo tempo, somente pesos de interesse são atualizados para alguns padrões de aprendizado, com derivada constante [16]; além disso a simplicidade computacional de implementação da não-linearidade, ao utilizar apenas a lógica condicional, ao invés de métodos numéricos iterativos ou a técnica *lookup table* para representar as funções de ativação transcendentais (e.g. tangente hiperbólica, sigmóide) [17].

A similaridade entre a função ReLU e a função Shrinkage, utilizada em [8], estimula a utilização da mesma como função de ativação do neurônio na camada de saída da RNA, substituindo a combinação linear, como havia sido realizado em [7] e representado na figura 5. Desta forma se combinam os métodos de RNA com representação esparsa dos dados. Resultando na estrutura de RNA representada na figura 7.

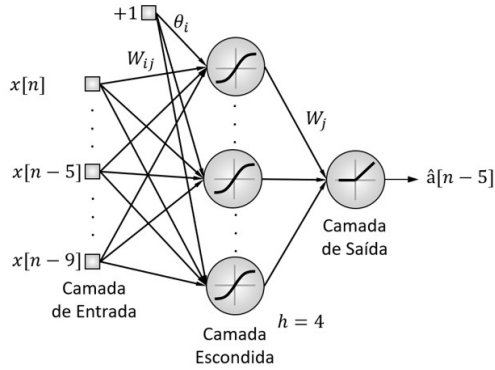


Fig. 7. Arquitetura da RNA com saída esparsa.

Motivando-se pela popularidade da função de ativação ReLU nas camadas ocultas [16], propõe-se a implementação da função ReLU em todos os neurônios da RNA, tornando tanto a saída esparsa, assim como a interconexão dos neurônios da RNA esparsa, conforme a figura 8.

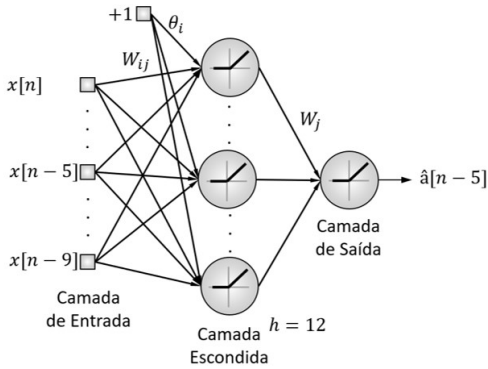


Fig. 8. Arquitetura da RNA esparsa.

Para comparação dos métodos definiu-se a métrica de erro como a raiz do erro quadrático médio (eRMS) entre o valor processado por cada estrutura proposta e a amplitude esperada de energia. Foram geradas 110.000 amostras no simulador de pares dos vetores de dimensão igual a quantidade de amostras $\{\mathbf{x}, \mathbf{a}\}$, sendo 10.000 amostras de pares utilizadas para treinamento dos modelos e 100.000 amostras para teste, especificando a capacidade de generalização dos modelos.

III. RESULTADOS

A quantidade de atrasos para entrada de todas RNAs foi mantida do trabalho desenvolvido por [7], por critério comparativo, considerando 10 atrasos da entrada. Devido o fato da função ReLU ser mais simples são necessários mais neurônios

com tal função na camada escondida da RNA. Logo, calculou-se o erro RMS para as quantidades de neurônios e ocupâncias do sinal de 10% a 90%, como apresentado na figura 9.

Pelo interesse de utilizar uma mesma arquitetura para todos os valores de ocupância a análise da escolha de neurônios foi reduzida para duas dimensões, medindo-se a média dos erros RMS de todas as ocupâncias para cada quantidade de neurônios, como apresentado na figura 10. Respeitando uma relação de simplicidade e desempenho, foi definida a arquitetura com 12 neurônios na camada oculta esboçada na figura 8.

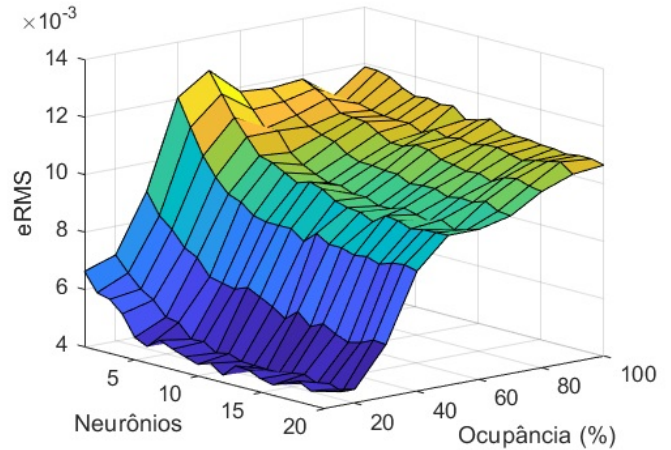


Fig. 9. Superfície de erro da arquitetura da RNA e Ocupância.

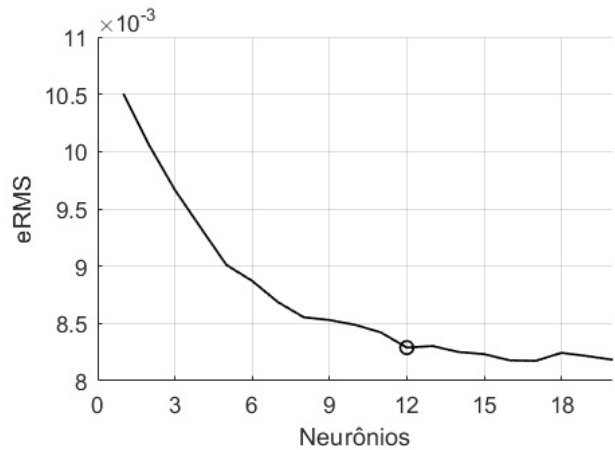


Fig. 10. Curva de erro para definição da quantidade de Neurônios.

Comparando-se a utilização do método FIR com representação esparsa (FIRShir), o método de RNA com ativação tangente hiperbólica (RNAtanh), o método proposto de combinação da RNAtanh com a representação esparsa (RNAtanhReLU) e a outra proposta de combinação da RNA esparsa em todos os neurônios (RNARELU), foram obtidos os resultados apresentados na figura 11, da qual percebe-se que a combinação dos métodos ocasionou desempenho melhor para todos os valores de ocupância em relação aos métodos de referência desse trabalho.

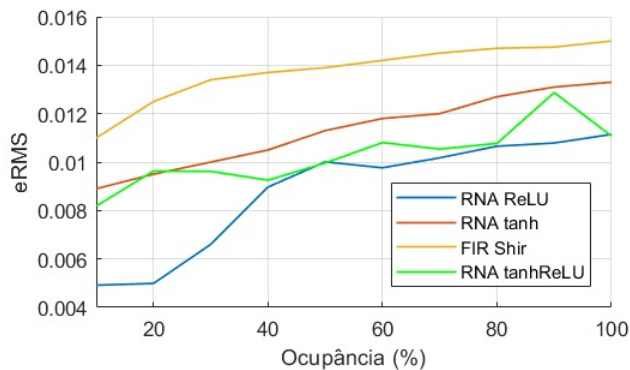


Fig. 11. Comparação de desempenho entre os estimadores.

IV. CONCLUSÕES

O aumento da taxa de eventos em cenários futuros previstos ao LHC demanda algoritmos de estimação de energia do Tilecal. Este trabalho propôs o uso de uma rede neural esparsa como uma alternativa para estimação de energia. Ao fim das análises, pode-se concluir que o estimador neural com função de ativação ReLU apresentou desempenho superior quando comparado a um método linear de representação esparsa e a RNAs com ativação tangente hiperbólica, visto a técnica de RNA esparsa coincidir com as especificações dos sinais do Tilecal. Além de se mostrar promissora na simplicidade de implementação do algoritmo. Para trabalhos futuros, objetiva-se a utilização de RNAs esparsas com aprendizado não-supervisionado, tornando a RNA independente do simulador.

AGRADECIMENTOS

Os autores são gratos ao CNPq, CAPES, FAPERJ, CERN (Suíça) e União Européia pelas contribuições para desenvolvimento dessa pesquisa.

REFERÊNCIAS

- [1] M. A. Moreira. O modelo padrão da física de partículas. Revista brasileira de ensino de física. São Paulo. Vol. 31, n. 1, 1306, 11 p., Março 2009.
- [2] D. A. Edwards e M. J. Syphers. *An introduction to the physics of high energy accelerators*. John Wiley & Sons, 2008.
- [3] L. M. de Andrade Filho, B. S. Peralva, J. M. de Seixas, and A. S. Cerqueira. *Calorimeter response deconvolution for energy estimation in high-luminosity conditions*. IEEE Transactions on Nuclear Science, 62(6):3265-3273, Dec 2015.
- [4] CERN. *The large hadron collider*, 2014.
- [5] ATLAS, C. *The atlas experiment at the cern large hadron collider*. Journal of Instrumentation, v. 3, n. 08, 2008, p. S08003.
- [6] F. Ruggiero. *LHC Accelerator R&D and Upgrade Scenarios*, The European Physical Journal C-Particles and Fields, v.34, 2004, pp.433-442.
- [7] M. H. M. Faria, L. M. de Andrade Filho, J. P. B. da S. Duarte e J. M. Seixas. Redes neurais para filtragem inversa com aplicação em calorímetros operando a alta taxa de eventos. XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBRT2017, 3-6 DE SETEMBRO DE 2017, São Pedro, SP, 2017.
- [8] J. P. B. da S. Duarte, L. M. de Andrade Filho e J. M. de Seixas. Representação Esparsa Aplicada à Estimação Online de Energia para um Calorímetro Submetido a uma Alta Taxa de Eventos. XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBRT2017, 3-6 DE SETEMBRO DE 2017, São Pedro, SP, 2017.
- [9] A. Henriques. *The atlas tile calorimeter*. In *Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA)*, 4th International Conference on 2015, IEEE, 2015.
- [10] B. S. Peralva. *The TileCal energy reconstruction for collision data using the matched filter*. In: Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 2013, p. 1-6.
- [11] K. Anderson et al. *Front-end Electronics for the ATLAS Tile Calorimeter*. Proceedings of Fourth Workshop on Electronics for LHC Experiments: Rome, 1998.
- [12] A. S. Cerqueira et al. *Tile Calorimeter Upgrade Program for the Luminosity Increasing at the LHC*. arXiv preprint arXiv:1509.08994, 2015.
- [13] J. Chapman. *ATLAS simulation computing performance and pile-up simulation in ATLAS*. In LPCC detector simulation workshop, CERN, 2011.
- [14] K. Burse, R. N. Yadav, S. Shrivastava. *Channel equalization using neural networks: A review*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), IEEE, v. 40, n. 3, 2010, p. 352-357.
- [15] S. HAYKIN. *Redes neurais: princípios e prática*, trad. Paulo Martins Engel. 2 ed. Porto Alegre: Bookman, 2001.
- [16] *Deep Learning Book*: Capítulo 8 - Funções de Ativação. Disponível em: <http://deeplearningbook.com.br/funcao-de-ativacao/>. Acesso em: 29 de Abril de 2019.
- [17] X. Glorot, A. Bordes e Y. Bengio. *Deep Sparse Rectifier Neural Networks*. Appearing in Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011, Fort Lauderdale, FL, USA. Volume 15 of JMLR: W&CP 15, 2011.