

# Avanços na Estimação da Energia do Calorímetro de Telhas do Experimento ATLAS no LHC

Gustavo B. Libotte \* Guilherme I. Gonçalves \* Bernardo S. Peralva \*  
Luciano M. de Andrade Filho \*\* José Manoel de Seixas \*\*\*

\* Instituto Politécnico, Universidade do Estado do Rio de Janeiro, RJ  
(e-mail: gustavolibotte@iprj.uerj.br, bernardo@iprj.uerj.br).

\*\* Faculdade de Engenharia, Universidade Federal de Juiz de Fora,  
MG (e-mail: luciano.andrade@ufjf.edu.br)

\*\*\* Laboratório de Processamento de Sinais, Universidade Federal do  
Rio de Janeiro, RJ (e-mail: seixas@lps.ufrj.br)

**Abstract:** This work explores performance improvements achieved by energy estimation methods typically used in high-energy calorimetry systems. The goal is to evaluate their operation in high-luminosity conditions, such as in the ATLAS experiment in the Large Hadron Collider (LHC). The high-luminosity conditions increases the number of signals produced at each collision in such a way that the electronics from the calorimeter systems acquire information from adjacent events, producing the signal pile-up effect which degrades the efficiency of the energy estimation methods. In this context, the Optimal Filter (OF2) technique, currently employed in the ATLAS hadronic calorimeter (TileCal), is evaluated. Unlike the current version of the OF2 used in TileCal, this work proposes the use of the signal pile-up information, through the noise covariance matrix, to compute the filter coefficients. The results show that the proposed improvement significantly increases the OF2 efficiency, especially for readout channels that are more sensitive to the signal pile-up effect.

**Resumo:** Este trabalho explora melhorias no desempenho de métodos de estimativa de energia tipicamente utilizados em calorímetros de altas energias. O objetivo é avaliar a operação destes métodos em condições de alta luminosidade, tais como as previstas no experimento ATLAS do Grande Colisor de Hádrons (LHC). A operação em alta luminosidade aumenta o número de sinais gerados a cada colisão, de modo que a eletrônica do sistema de calorimetria adquire sinais de colisões adjacentes, produzindo o efeito de empilhamento de sinais o qual degrada a eficiência dos métodos de estimação da energia. Neste contexto, este estudo concentra-se no método do Filtro Ótimo (OF2) atualmente em operação no principal calorímetro hadrônico do experimento ATLAS no LHC, o TileCal. Diferentemente da versão atual do método OF2 empregada no TileCal, este trabalho propõe o uso da informação do empilhamento de sinais, através da matriz de covariância do ruído, para o cálculo dos coeficientes do filtro. Os resultados mostram que o aperfeiçoamento proposto neste trabalho aumenta expressivamente a eficiência do método OF2, especialmente para canais de leitura com alta incidência de partículas.

**Keywords:** ATLAS experiment; High-Luminosity LHC; Energy estimation; Signal pile-up effect, High-energy calorimetry.

**Palavras-chaves:** Experimento ATLAS; LHC de Alta Luminosidade; Estimação de energia; Efeito de empilhamento de sinais; Calorimetria de altas energias.

## 1. INTRODUÇÃO

Experimentos de Física de Alta Energia geralmente exigem sistemas de instrumentação sofisticados, especialmente quando colisores de partículas estão envolvidos. Para analisar fenômenos físicos cada vez mais raros, há uma tendência de aumentar as taxas de colisão e o número de partículas por colisão, o que aumenta a luminosidade alcançada pelo experimento. Este é o caso do LHC (Large

Hadron Collider) no CERN (Organização Europeia para a Pesquisa Nuclear), o colisor de partículas mais avançado já construído. O LHC impulsiona prótons a uma energia de 6,5 TeV e os colide em quatro pontos de interação, com uma frequência de 40 MHz. Instalado em um desses pontos está o detector ATLAS, que registra as propriedades das partículas produzidas nas colisões para posterior análise física. O experimento ATLAS investiga uma ampla gama de tópicos de física, com foco principal em melhorar a compreensão dos constituintes fundamentais da matéria e das interações entre eles (Sotto-Maior Peralva, 2015; The ATLAS Collaboration, 2020; Teixeira et al., 2021).

\* GBL e BSP agradecem à Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) pelo apoio financeiro, através dos processos E-26/210.430/2024 e E-26/201.304/2022, respectivamente.

O LHC está constantemente aumentando o seu desempenho para possibilitar mais descobertas científicas. O objetivo é alcançar luminosidades instantâneas de 5 a 7,5 vezes maiores do que o valor nominal do LHC, permitindo assim que os experimentos ampliem suas amostras de dados em uma ordem de magnitude durante os 12 anos de operação da versão atualizada do LHC, chamada de HL-LHC (High-Luminosity Large Hadron Collider). A luminosidade é crucial, pois está diretamente ligada ao número de colisões que ocorrem em um determinado período de tempo, permitindo a observação de processos raros. O HL-LHC, previsto para operar a partir de 2029, possibilitará aos físicos estudar mecanismos conhecidos, como o bóson de Higgs, com maior detalhe, além de observar novos fenômenos raros. Por exemplo, o HL-LHC deverá produzir pelo menos 15 milhões de bósons de Higgs por ano, em comparação com os cerca de três milhões produzidos pelo LHC em 2017 (Gómez Delegido, 2023; Leone, 2024).

O HL-LHC está previsto para fornecer uma luminosidade integrada de  $3.000 \text{ fb}^{-1}$ , o que representa dez vezes mais dados do que foi coletado até 2023. Essa alta estatística possibilitará realizar medidas mais precisas e aprimorar buscas por nova física. Prevê-se que a luminosidade máxima seja de  $7,5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , correspondendo a cerca de 200 interações próton-próton por evento, o que aumentará a demanda de processamento na etapa de seleção de eventos (*trigger*) e degradará o desempenho de estimativas de energia, em virtude da amplificação do efeito de empilhamento de sinais. Isso representa um desafio significativo para a reconstrução de trajetórias e os requisitos computacionais associados devido ao número sem precedentes de eventos por intervalo de tempo. Para enfrentar esse desafio, é fundamental o aprimoramento de algoritmos dedicados visando mitigar o efeito e otimizar ainda mais os algoritmos padrão do detector ATLAS (Testa and The ATLAS Collaboration, 2019).

Neste contexto, o objetivo deste trabalho é apresentar uma análise de desempenho do método OF2 (Optimal Filter) (Fullana et al., 2006), atualmente empregado no principal calorímetro hadrônico do ATLAS, o TileCal, e propor melhorias no projeto deste estimador. A versão atual utilizada no TileCal não prevê o uso da informação de empilhamento de sinais no projeto do filtro. Logo, a eficiência do método OF2 é deteriorada em condições de alto empilhamento de sinais. Desta forma, este trabalho busca comparar a eficiência da estratégia atualmente em operação com a eficiência da versão otimizada proposta neste trabalho, visando avaliar o ganho no processo de estimativa de energia em um cenário de alta luminosidade. Para isso, conduzimos análises com dados de Monte Carlo de viés mínimo, que simulam eventos capturados durante uma tomada de dados no acelerador de partículas sob condições específicas. As análises se dão tanto a nível de célula quanto a nível de canal de leitura no calorímetro, com o propósito de mitigar o efeito de empilhamento de sinais durante o processo de reconstrução de energia do TileCal.

Na próxima seção, o ambiente de calorimetria de altas energias é apresentado assim como o problema de empilhamento de sinais. A Seção 3 descreve o método OF2 e introduz a contribuição deste trabalho visando o aperfeiçoamento deste método. A base de dados utilizada e os

resultados alcançados são mostrados na Seção 4. Por fim, as conclusões são derivadas na Seção 5.

## 2. O SISTEMA DE CALORIMETRIA DE ALTAS ENERGIAS DO ATLAS

O experimento ATLAS está fortemente associado ao seu sistema de calorimetria de ponta para absorver e medir a energia das partículas produzidas em colisões. Os dados dos calorímetros são cruciais para determinar as propriedades-chave das partículas, desempenhando um papel fundamental na diferenciação entre os tipos de partículas. O sistema de calorimetria do ATLAS inclui dois componentes principais: o Calorímetro de Argônio Líquido (LAr) para calorimetria eletromagnética e o Calorímetro Hadrônico de Telhas (TileCal) para medidas de energia hadrônica (The ATLAS Collaboration, 2010; Wigmans, 2017), onde o segundo é elemento central neste estudo.

O TileCal é o principal calorímetro hadrônico do experimento ATLAS no LHC. Do ponto de vista da construção, o TileCal é um detector de amostragem feito de placas de aço como absorvedor e telhas cintilantes como meio ativo. O TileCal absorve energia das partículas, principalmente hárds, que passam pelo calorímetro eletromagnético do ATLAS. Fisicamente, o TileCal é composto por três partes cilíndricas: um barril central de telhas logicamente dividido em duas partições (Long Barrel A e Long Barrel C, ora referidas como LBA e LBC, respectivamente) e dois barris estendidos de telhas (Extended Barrel A e Extended Barrel C, referenciadas como EBA e EBC, respectivamente). Cada parte é composta por 64 módulos para construir o cilindro inteiro. O módulo do barril central e ambos os barris estendidos são ainda divididos, respectivamente, em 45 e 16 células com dupla leitura. Cobrindo uma ampla faixa de pseudo-rapidez de  $|\eta| < 1,7$  com cobertura azimutal completa (Peralva and ATLAS Tile Calorimeter System, 2022). A Figura 1 ilustra o sistema de calorimetria do ATLAS.

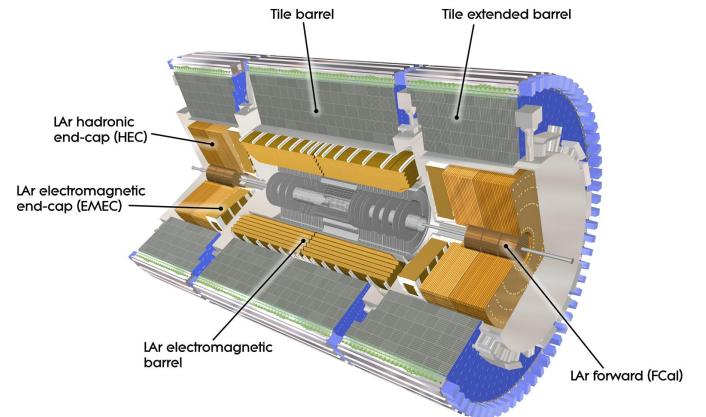


Figura 1. Esquema do detector interno e calorímetros do ATLAS, onde o TileCal é formado por dois barris centrais e outras duas seções estendidas, envolvendo os calorímetros eletromagnéticos LAr (Marjanovic, 2019).

A luz produzida por cada telha cintilante é coletada por fibras (WLS, *wavelength shifting*) e transmitida a tubos fotomultiplicadores (PMT, *photo-multiplier tubes*), onde

um sinal elétrico é produzido, compondo cada canal de leitura. Este sinal elétrico é posteriormente condicionado na eletrônica no detector por meio de um circuito de formação de sinais antes de ser digitalizado a uma frequência de amostragem de 40 MHz com conversores analógico-digitais (ADCs) de 10 bits (Anderson et al., 2005; Jen-La Plante and Tylmad, 2010).

O TileCal é uma estrutura cilíndrica com um raio interno de 2,28 m e um raio externo de 4,23 m, onde o Barril Central (LB) cobre  $|\eta| < 1$  e o Barril Estendido (EB) cobre  $0,8 < |\eta| < 1,7$ , onde  $\eta$  é a pseudorapidez, que essencialmente representa a posição angular em relação ao eixo do feixe. Cada barril consiste em 64 módulos em forma de cunha dispostos na direção  $\phi$ . Os módulos no TileCal são segmentados em três camadas de células, A, BC e D, com granularidade de célula de  $0,1 \times 0,1$  para as camadas A e BC e  $0,2 \times 0,1$  para a camada D, como mostrado na Figura 3. Além disso, o TileCal possui uma camada especial de células (camada E) composta apenas de cintiladores sem material pesado. Cada módulo nos barris central e estendido contém, respectivamente, 23 e 16 células de dupla leitura, totalizando aproximadamente 10.000 canais de leitura (The ATLAS Collaboration, 2008).

## 2.1 O problema de empilhamento de sinais

Para otimizar o potencial de novas medições e descobertas no LHC, é fundamental maximizar as taxas de colisão e, como resultado direto, ocorrem múltiplas colisões próton-próton toda vez que dois feixes de prótons se intersectam. Colisões de alta energia de interesse podem, portanto, ser frequentemente ofuscadas por várias colisões de baixa energia e sem viés direcional. Esse fenômeno, conhecido como empilhamento de sinais, contamina o estado final da colisão, tornando a reconstrução dos objetos neste estado final uma tarefa desafiadora e resultando em um aumento das incertezas nas medidas experimentais (Soyez, 2019).

No TileCal, colisões consecutivas dentro da mesma janela de leitura são propiciadas, especialmente em canais com alta ocupação, pois a forma do pulso se estende além do espaçamento do feixe do LHC, resultando em um intervalo de tempo mais amplo para a leitura (aproximadamente 150 ns). Embora, teoricamente, uma única amostra no pico máximo do sinal deveria ser suficiente, na prática, variações temporais introduzem imprecisões adicionais nas medições de energia, afetando a amplitude do pulso de forma proporcional.

Devido ao tempo de decaimento prolongado, múltiplos pulsos de colisões de feixes consecutivos podem se sobrepor. Esse problema é intensificado pelas colisões próton-próton simultâneas no mesmo ponto de cruzamento de feixes, resultando no que é conhecido como empilhamento de sinais no tempo. Essa distorção compromete o sinal desejado e prejudica a precisão da estimativa de amplitude, pois emerge quando a resposta do calorímetro não consegue acompanhar o ritmo das colisões. Ele contribui para a resposta do sinal e para a ampliação da distribuição de energia da célula, aumentando com o número médio de interações por passagem de feixe  $\langle\mu\rangle$  (Klimek and The ATLAS Collaboration, 2020; Gonçalves and ATLAS Tile Calorimeter Group, 2022).

O fenômeno de empilhamento de sinais é comum em experimentos de alta taxa de eventos, onde a leitura de um evento é contaminada por eventos adjacentes devido à sobrecarga na capacidade de processamento do detector. Para se ter uma idéia, operação atual do LHC (*Run 3*) atinge uma média de 50 a 80 interações por colisão  $\langle\mu\rangle$ , que acontece a cada 25 ns. Após a fase de atualização do LHC, entre 2024 e 2026, espera-se que o LHC alcance valores ainda mais elevados, possivelmente até 200 interações por colisão durante as campanhas de aquisição de dados das *Runs 4* e *5* (2026 a 2038) (Apollonio et al., 2023).

## 3. ESTRATÉGIA PARA RECONSTRUÇÃO DE SINAIS

O método OF2 (Fullana et al., 2006), que atualmente está em operação no TileCal, modela as amostras de tempo recebidas  $\mathbf{x}$  no instante  $k$  como uma função linear do pulso de referência, do ruído e do pedestal, representados por  $A g_{k-\tau}$ ,  $n_k$ , e  $\text{ped}$ , respectivamente, de tal modo que

$$x_k = A g_{k-\tau} + n_k + \text{ped}, \quad (1)$$

onde  $A$  denota a amplitude do pulso,  $g_k$  é o vetor contendo as amostras de tempo normalizadas da forma do pulso de referência, e  $\tau$  é a defasagem de fase.

Em condições de baixa luminosidade, onde o efeito do empilhamento pode ser negligenciado sem prejuízos, o ruído é predominantemente eletrônico e estacionário, o que permite o uso de técnicas de minimização de variância, sujeita a restrições de não tendenciosidade, robustez e linearidade, para estimar o parâmetro de interesse, a amplitude do sinal  $A$ . O algoritmo OF2 utiliza essa condição para desenvolver um estimador linear baseado em uma operação de soma ponderada, onde a estimativa da amplitude do sinal,  $\hat{A}$ , é recuperada através de uma operação de produto escalar entre os pesos  $\mathbf{w}$  e o vetor de amostras de tempo  $\mathbf{x}$ , isto é,

$$\hat{A} = \sum_{k=1}^N x_k w_k, \quad (2)$$

com  $w_k$  denota os pesos do OF2 correspondentes a  $N$  amostras.

A linearidade do modelo é assegurada pela aproximação do sinal recebido como a expansão de Taylor de primeira ordem

$$x_k = A g_k - A \tau \dot{g}_k + n_k + \text{ped}, \quad (3)$$

para  $k = 1, 2, \dots, N$ , onde  $\dot{g}_k$  representa a derivada de  $g_k$  no tempo.

Para garantir uma estimativa não tendenciosa, três restrições são impostas aos pesos do OF2,

$$\sum_{k=1}^N w_k g_k = 1, \quad \sum_{k=1}^N w_k \dot{g}_k = 0, \quad \text{e} \quad \sum_{k=1}^N w_k = 0. \quad (4)$$

Essas restrições garantem que a estimativa seja insensível a flutuações na linha de base do pulso e desvios na forma do pulso devido a variações de fase.

A minimização da variância, que é a medida de dispersão da estimativa em torno do valor verdadeiro da amplitude do sinal, é realizada sob as restrições mencionadas acima.

Isso é alcançado através de multiplicadores de Lagrange  $\lambda$ ,  $\xi$  e  $\nu$ , conforme mostrado na expressão

$$I_w = \sum_{k,j=1}^N w_k w_j C_{k,j} - \lambda \left( \sum_{k=1}^N w_k g_k \right) - \xi \left( \sum_{k=1}^N w_k \dot{g}_k \right) - \nu \left( \sum_{k=1}^N w_k \right). \quad (5)$$

A obtenção dos pontos críticos da Eq. (5), visando a minimização da variância, leva ao sistema de equações lineares dado por

$$\begin{pmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,N} & -g_1 & -\dot{g}_1 & -1 \\ C_{2,1} & C_{2,2} & \cdots & C_{2,N} & -g_2 & -\dot{g}_2 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ C_{N,1} & C_{N,2} & \cdots & C_{N,N} & -g_N & -\dot{g}_N & -1 \\ g_1 & g_2 & \cdots & g_N & 0 & 0 & 0 \\ \dot{g}_1 & \dot{g}_2 & \cdots & \dot{g}_N & 0 & 0 & 0 \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \lambda \\ \xi \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

A solução para os pesos do OF2 é obtida resolvendo-se este sistema de equações lineares, onde a matriz de correlação de ruído  $\mathbf{C}$  desempenha um papel crucial. Embora em condições ideais possa ser aproximada como a matriz identidade, em cenários mais realistas, como em condições de empilhamento, a matriz  $\mathbf{C}$  precisa ser estimada para levar em conta correlações e não linearidades introduzidas pelo empilhamento do sinal, uma vez que a matriz de covariância do ruído absorve as informações estatísticas de segunda ordem do empilhamento de sinal (Barbosa et al., 2017).

Em suma, a matriz de covariância do ruído desempenha um papel crucial no projeto do OF2, especialmente em cenários de alta luminosidade, onde a distribuição do ruído é afetada por estatísticas de segunda ordem introduzidas pelo empilhamento de sinal. O uso adequado dessas estatísticas é essencial para garantir um desempenho ótimo do método OF2 sob diferentes condições de operação do calorímetro. Portanto, ao considerar a matriz de correlação de ruído correta, o desempenho do método OF2 pode ser significativamente aprimorado. Essa abordagem robusta e eficaz do OF2 é crucial para garantir reconstruções mais precisas da energia do sinal no TileCal. Sendo assim, a principal contribuição deste trabalho se refere na comparação das eficiências das versões otimizada e não otimizada do método OF2.

Daqui para frente, vamos nos referir ao método que toma  $\mathbf{C}$  como sendo a matriz de covariância como OF2 COV e vamos comparar o seu desempenho com o método OF2 típico, onde a matriz  $\mathbf{C}$  é igual à matriz identidade de dimensão equivalente. Note que, na literatura, este método também pode ser designado apenas como OF, no caso em que a restrição associada ao pedestal na Eq. (4) é removida.

## 4. RESULTADOS

### 4.1 Base de dados

Com o objetivo de avaliar o desempenho da versão atualizada do método OF2 proposto, e comparar com a eficiência

da versão atualmente empregada no TileCal, foi utilizado um conjunto de dados de Monte Carlo produzido pela colaboração ATLAS nas condições de luminosidade do LHC próximas às atuais, ou seja,  $\langle\mu\rangle = 55$ . O conjunto, na sua totalidade, compreende 200.000 sinais contendo somente informação de empilhamento de sinais (ruído, neste caso) de modo que os métodos de estimativa devem ser projetados para reduzirem, ao máximo, a contribuição da energia destes eventos.

### 4.2 Análise de eficiência

Uma forma de avaliar a eficiência de um estimador imparcial é quantificar a sua variância. Deste modo, estimadores com valores pequenos de variância apresentam melhor eficiência quando comparados a estimadores de alta variância. Visto que a base de dados utilizada apresenta somente eventos de empilhamento de sinais, a estimativa da energia destes sinais resulta em uma tarefa desafiadora devido à sobreposição de múltiplos sinais. No entanto, ao aplicar o método OF2, particularmente com o uso de matrizes de covariância do ruído, conseguimos não apenas mitigar os efeitos adversos do empilhamento de sinais, mas também alcançar estimativas de energia mais precisas. Isso sugere que o OF2 pode ser uma ferramenta poderosa para lidar com eventos de empilhamento de sinais, oferecendo estimativas de energia com menor variância e, portanto, maior eficiência. Essa abordagem não só melhora a qualidade dos dados reconstruídos, mas também fortalece a capacidade do experimento em extrair informações significativas sobre as propriedades das partículas envolvidas nas colisões.

Tipicamente, a medida de desempenho adotada no TileCal para estimar a eficiência de uma reconstrução é a raiz quadrada do erro médio (RMS) entre a energia (ou amplitude) estimada e de referência. Em particular, para os dados adotados nesta análise, a energia de referência é igual a zero, uma vez que os dados contêm apenas informação de ruído. A Figura 2 mostra histogramas de reconstrução de energia usando os métodos OF2 (em cinza) e OF2 COV (em vermelho) para dados relativos à célula A13 do calorímetro.

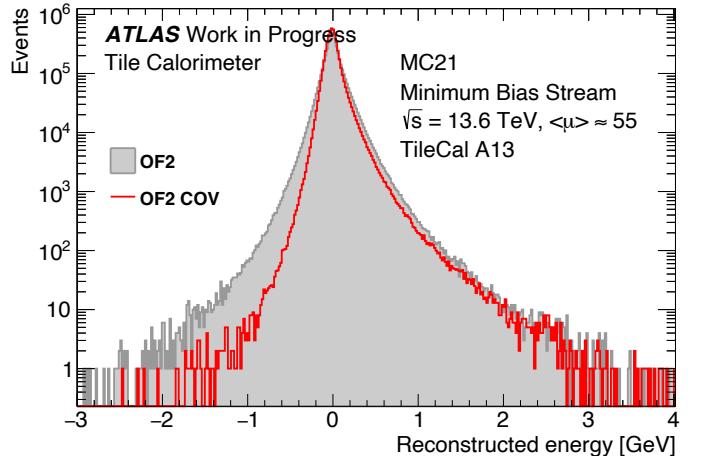


Figura 2. Histogramas de energia reconstruída utilizando dados de empilhamento de sinais para a célula A13 usando ambas as abordagens.

De forma geral, células da camada A estão mais próximas do feixe de partículas. Isso significa que a probabilidade

de capturarem eventos adjacentes em intervalos de tempo menores do que o tempo mínimo entre dos sinais completos é maior. Com isso, estas células tendem a apresentar particularmente maior ocupação, o que está diretamente ligado ao efeito do empilhamento de sinais. A Figura 2 deixa evidente este fato: ao adotar-se a matriz de covariância do ruído (OF2 COV) no procedimento de geração dos pesos e, consequentemente, na reconstrução de energia, observa-se um histograma com menor dispersão, que manifesta-se principalmente na calda negativa em comparação ao resultado correspondente do OF2. Isso traduz-se em um menor valor de RMS e, portanto, numa reconstrução mais eficiente.

A seguir, avaliamos o desempenho comparativo dos métodos analisados no domínio das células e canais de leitura do TileCal, levando em conta os resultados apresentados na Figura 3 para um nível de luminosidade  $40 < \mu < 60$ . Nestes resultados é considerada a mesma métrica discutida na Figura 2, o RMS do erro de estimativa. Inicialmente, considere a análise a nível de célula, apresentada na Figura 3(a). Neste esquema, está representado o arranjo espacial das células dos barris LBA e EBA, dispostas em relação ao eixo horizontal do feixe de partículas e a pseudorapidez  $\eta$ . As cores associadas a cada célula estão relacionadas à diferença relativa percentual entre os valores de RMS do erro de estimativa obtidos para cada célula pelos métodos comparados. Neste caso, seja  $\varepsilon_c^m$  o valor do RMS do erro de estimativa calculado para a célula  $c$ , usando um método  $m \in \{\text{OF2 COV}, \text{OF2}\}$ . A diferença relativa  $\mathcal{D}_c$  é dada por

$$\mathcal{D}_c = \frac{\varepsilon_c^{\text{OF2 COV}} - \varepsilon_c^{\text{OF2}}}{\varepsilon_c^{\text{OF2}}} \cdot 100\% \quad (6)$$

Assim, se  $\varepsilon_c^{\text{OF2 COV}} > \varepsilon_c^{\text{OF2}}$ , então  $\mathcal{D}_c > 0$ . Esta condição decorre do método OF2 ter desempenho superior na reconstrução de energia na célula  $c$ . Por outro lado, o método OF2 COV demonstra desempenho superior quando  $\mathcal{D}_c < 0$ .

De acordo com a escala de cores para a diferença relativa percentual na Figura 3(a), vemos que em todas as células o método OF2 COV apresenta desempenho superior, haja vista que  $\mathcal{D}_c < 0$  para todo  $c$ . As células da camada A apresentam valores negativos de diferença relativa com a maior ordem de grandeza, com diferenças de desempenho entre os métodos que alcançam mais de 25%. Como as células desta camada estão mais próximas do feixe de partículas, seus canais de leitura capturam mais eventos do subproduto da colisão entre partículas, em comparação com as células das demais camadas. Isso faz com que estas células apresentem maior ocupação dos seus canais de leitura em intervalos de tempo arbitrários, intensificando o efeito do empilhamento de sinais. Assim, a matriz de covariância captura estatísticas de segunda ordem, que favorecem desempenho superior do método OF2 COV.

Canais de leitura das células de camadas superiores, BC e D, têm tendência de menor ocupação uma vez que capturam eventos não registrados na camada A, que está mais exposta ao feixe de partículas. À medida que as colisões se sucedem, o feixe se torna menos denso, produzindo proporcionalmente menos eventos pela colisão entre partículas. Com isso, o efeito do empilhamento de sinais tende

a ser gradativamente menor nas células de camadas mais distantes do feixe, de tal modo que o método OF2 COV passa a ter desempenho menos destacado em relação ao OF2. Em condições de baixa ocupação, o empilhamento de sinais pode ser visto como *outlier*, promovendo efeito menos relevante na matriz de covariância, o que se traduz em desempenho mais próximo entre os métodos.

Canais de leitura do TileCal com ângulos de incidência próximos ao eixo do acelerador (altos valores de  $\eta$ ) são particularmente suscetíveis ao efeito de empilhamento de sinais, devido à sua maior exposição aos subprodutos das colisões. Este fato deixa claro a razão para células posicionadas entre  $0,9 < \eta < 1,6$  terem valores negativos de diferença relativa com a maior ordem de grandeza, em relação à sua respectiva camada. Destacadamente, vemos que os canais A12–A15 têm valores negativos de diferença relativa com a maior ordem de grandeza, pois são as células onde o efeito do empilhamento de sinais é mais severo e, portanto, o emprego das matrizes de covariância no cálculo dos pesos, em detrimento da matriz identidade, pode oferecer a maior vantagem.

Agora considere a análise de desempenho a nível de canal de leitura, como mostrado na Figura 3(b). O desempenho é avaliado para cada canal instrumentado em todos os 64 módulos no LBA, atribuindo-se a cada canal um valor de diferença relativa percentual, onde  $c$  denota o canal de leitura na Eq. (6) neste caso. Neste caso, canais associados à cor branca são aqueles não instrumentados. Neste contexto, podemos ver que a Figura 3(b) é predominantemente composta por canais em que a diferença relativa percentual é negativa.

Visando complementar a compreensão acerca destes resultados, a Figura 3(c) apresenta a frequência de canais em que o valor da diferença relativa percentual está dentro de intervalos arbitrários. Isso permite quantificar canais em que cada um dos métodos apresentam diferentes níveis de desempenho, complementando a análise visual da Figura 3(b). Dos 2.880 canais instrumentados avaliados no LBA, apenas em cerca de 10% o método OF2 teve desempenho superior ao OF2 COV, sendo que na maioria destes canais a diferença não ultrapassou 6%. A maior contribuição para estes resultados advém das células D0–D2, que estão associadas as canais 0, 13, 14, 24 e 25, como podemos perceber nas faixas correspondentes em tom amarelo na Figura 3(b). Estes resultados concordam com as análises a nível de célula, considerando as mesmas razões discutidas no contexto da Figura 3(a).

Por outro lado, há percentuais significativos de canais com diferença relativa percentual entre  $[-24, 0]$ , em cada faixa compreendida neste intervalo, tal que estes canais somam para mais de 90% do total de canais avaliados. Isso demonstra que, de forma geral, o empilhamento de sinais exerce papel muito relevante durante o processo de estimativa de energia. Ao considerar a matriz de covariância do ruído, o conjunto de pesos corrigidos fornece estimativas de energia que compensam as distorções de sinal ocasionadas pela leitura sobreposta de sinais adjacentes. Como estas informações são negligenciadas ao considerar-se  $\mathbf{C}$  como a matriz identidade na Seção 3, a tendência é que o ganho em se usar o OF2 COV seja amplificado à medida que o nível de luminosidade aumente.

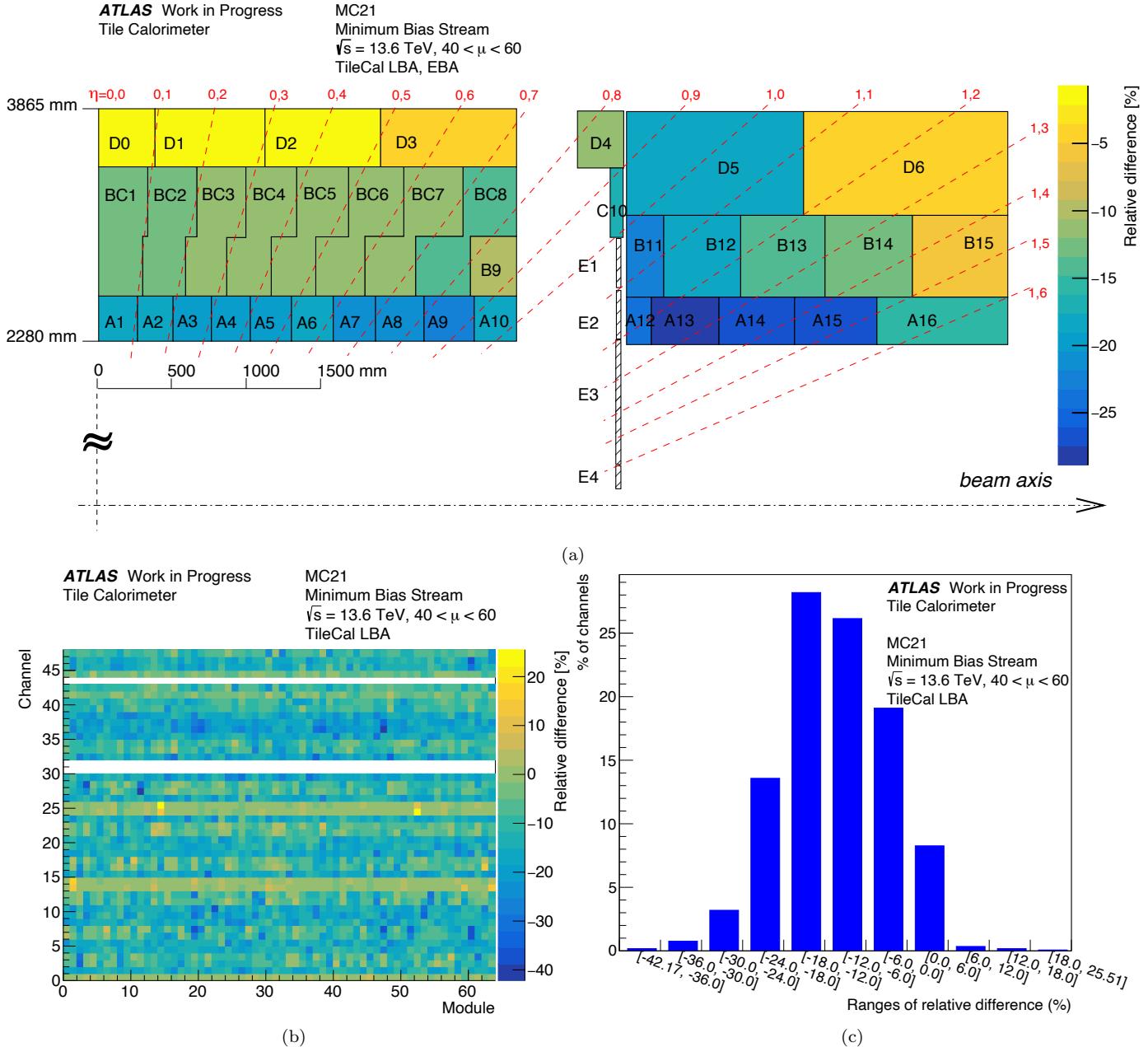


Figura 3. Diferença relativa percentual do RMS da energia reconstruída por cada um dos métodos, a nível de (a) célula e (b) canal de leitura; (c) frequência de canais com valores de diferença relativa percentual para determinadas faixas de desempenho. Nestes resultados, diferenças positivas representam desempenho superior do método OF2, enquanto que diferenças negativas refletem o ganho do OF2 COV.

Agora considere uma análise mais ampla, onde todas as partições do calorímetro são levadas em conta, de modo que o desempenho seja avaliado em termos do RMS do erro de estimativa para as células em função do valor de  $\eta$ , como mostrado na Figura 4. Novamente, consideramos níveis de luminosidade onde  $40 < \mu < 60$ , concordando com os resultados da Figura 3. No quadro superior, o desempenho está dividido de acordo com as camadas do calorímetro, respeitando o esquema de cores disposto na legenda. Marcadores preenchidos estão associados ao resultados obtidos com o OF2, enquanto que marcadores delineados representam os resultados do OF2 COV. As barras de erro quantificam o desvio-padrão em torno dos resultados médios obtidos para  $k = 10$  dobradas com

a técnica de validação cruzada  $k$ -fold (Wong and Yeh, 2020). Por sua vez, no quadro inferior desta figura, temos a razão do RMS do erro de estimativa entre ambos os métodos, para resultados com correspondência de  $\eta$  na mesma camada.

De forma geral, estes resultados demonstram desempenho superior do OF2 COV na reconstrução de sinais em condições de alta luminosidade. Em conformidade com os resultados anteriores, diferenças mais expressivas do desempenho entre os métodos são observadas nas camadas mais próximas ao feixe de partículas. Na camada A, por exemplo, a razão do RMS do erro de estimativa entre ambas as abordagens alcança níveis abaixo de 80%, o que pode ser interpretado como uma melhoria proporcional

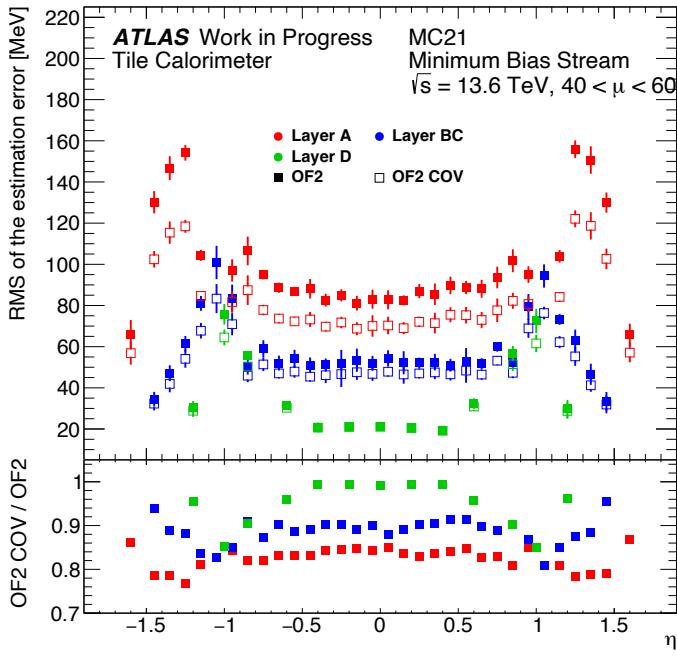


Figura 4. Comparativo do erro de estimativa utilizando as duas versões do método OF2 avaliadas em função de  $\eta$ . O quadro superior apresenta o RMS do erro de estimativa para diferentes camadas, destacando resultados de OF2 (marcadores preenchidos) e OF2 COV (marcadores delineados), com barras de erro representando desvios-padrão para  $k = 10$  dobras de validação cruzada. O quadro inferior exibe a razão do RMS entre os métodos para correspondentes valores de  $\eta$  na mesma camada.

a mais de 20% nas regiões com ângulos de incidência próximos ao eixo do acelerador. Mesmo nas regiões mais próximas a  $\eta = 0$ , a razão entre resultados correspondentes está próxima a 80%, o que ainda exprime um ganho considerável.

No caso da camada D, a maior diferença se dá para  $\eta \approx \pm 1$ , que refere-se à célula D5. Este resultado concorda com os valores de diferença relativa obtidos na Figura 3(a). Para os demais valores de  $\eta$  nesta camada, o desempenho do OF2 COV foi, pelo menos, igual ao do OF2, tendo em vista a razão entre os resultados próxima a um. Neste caso, a matriz de covariância não tem efeito preponderante, a ponto de levar a um ganho mais expressivo, uma vez que o efeito de empilhamento tende a ser menos influente nesta camada. Além disso, a magnitude das barras de erro revela níveis de confiança razoáveis para os resultados obtidos nas condições impostas.

## 5. CONCLUSÕES

Os resultados obtidos neste estudo proporcionam *insights* sobre o desempenho do método OF2 e sua versão aprimorada OF2 COV na reconstrução de energia em condições de alta luminosidade no calorímetro TileCal. Nossa análise revelou uma melhoria significativa no desempenho da reconstrução de energia ao incorporar informações sobre a covariância do ruído, evidenciando a eficácia do método OF2 COV em mitigar os efeitos adversos do empilhamento de sinais e fornecer estimativas mais precisas de energia.

Os histogramas de energia reconstruída demonstraram uma redução na dispersão dos resultados quando se utiliza o método OF2 COV em comparação com o OF2 padrão, especialmente em células próximas ao feixe de partículas, onde o empilhamento de sinais é mais pronunciado. Além disso, a análise por região e camada do calorímetro destacou o benefício do uso da matriz de covariância em células próximas ao feixe de partículas e em regiões com alta densidade de eventos, enquanto nas camadas mais distantes do feixe, onde o empilhamento de sinais é menos influente, o desempenho entre os métodos tende a se igualar. A análise em função do ângulo de incidência  $\eta$  ressaltou a importância de uma cobertura angular abrangente para a eficácia do método, evidenciando que o ganho de desempenho do OF2 COV é mais significativo em regiões com ângulos próximos ao eixo do acelerador.

Portanto, o método OF2 COV representa uma alternativa simples e eficiente para a reconstrução de energia em ambientes de alta luminosidade, como os encontrados no LHC, especialmente após o *upgrade* da eletrônica, que ampliará substancialmente a capacidade do calorímetro. Sua capacidade de melhorar significativamente a qualidade dos dados reconstruídos e fortalecer a capacidade dos experimentos em extraer informações relevantes sobre as propriedades das partículas envolvidas nas colisões sugere um potencial significativo para avanços futuros na área da física de altas energias.

## AGRADECIMENTOS

Os autores agradecem a CAPES, CNPq, FAPERJ, FAPEMIG e RENAFAE pelo suporte. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Código de Financiamento 001.

## REFERÊNCIAS

- Anderson, K., Gupta, A., Merritt, F., Oreglia, M., Pilcher, J., Sanders, H., Shochet, M., Tang, F., Teuscher, R., Wu, H., Blanchot, G., Cavalli-Sforza, M., and Korolkov, I. (2005). Design of the front-end analog electronics for the ATLAS tile calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 551(2-3), 469–476. doi:10.1016/j.nima.2005.06.048.
- Apollonio, A., Buffat, X., Bruce, R., De Maria, R., Giovanazzi, M., Iadarola, G., Lechner, A., Métral, E., Sterbini, G., Tomás, R., and Zerlauth, M. (2023). HL-LHC configuration and operational challenges. In *The Future of the Large Hadron Collider*, chapter 6, 95–120. WORLD SCIENTIFIC. doi:10.1142/9789811280184\_0006.
- Barbosa, D.P., Andrade Filho, L.M.A., Peralva, B.S., Cerqueira, A.S., and Seixas, J.M. (2017). Sparse Representation for Signal Reconstruction in Calorimeters Operating in High Luminosity. *IEEE Transactions on Nuclear Science*, 64(7), 1942–1949. doi:10.1109/TNS.2017.2712420.
- Fullana, E., Castelo, J., Castillo, V., Cuenca, C., Ferrer, A., Higón, E., Iglesias, C., Munar, A., Poveda, J., Ruiz-Martinez, A., Salvachua, B., Solans, C., Teuscher, R., and Valls, J. (2006). Digital Signal Reconstruction in the

- ATLAS Hadronic Tile Calorimeter. *IEEE Transactions on Nuclear Science*, 53(4), 2139–2143. doi:10.1109/TNS.2006.877267.
- Gómez Delegido, A.J. (2023). Upgrade of ATLAS Hadronic Tile Calorimeter for the High Luminosity LHC. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1047, 167860. doi:10.1016/j.nima.2022.167860.
- Gonçalves, G.I. and ATLAS Tile Calorimeter Group (2022). Energy Reconstruction Techniques in TileCal Under High Pile-Up Conditions. In *Communications in Computer and Information Science*, 140–151. doi:10.1007/978-3-030-96878-6\_13.
- Jen-La Plante, I. and Tylmad, M. (2010). Pulse shapes for signal reconstruction in the ATLAS Tile Calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 617(1-3), 96–98. doi:10.1016/j.nima.2009.09.114.
- Klimek, P. and The ATLAS Collaboration (2020). Performance of the ATLAS Tile Calorimeter in LHC Run-2. *Journal of Instrumentation*, 15(04), C04034–C04034. doi:10.1088/1748-0221/15/04/C04034.
- Leone, S. (2024). Upgrade of ATLAS hadronic Tile Calorimeter for the High Luminosity LHC. *Journal of Instrumentation*, 19(01), C01036. doi:10.1088/1748-0221/19/01/C01036.
- Marjanovic, M. (2019). ATLAS Tile Calorimeter Calibration and Monitoring Systems. *IEEE Transactions on Nuclear Science*, 66(7), 1228–1235. doi:10.1109/TNS.2019.2921941.
- Peralva, B.S.M. and ATLAS Tile Calorimeter System (2022). Upgrade of ATLAS hadronic Tile Calorimeter for the High-Luminosity LHC. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1038, 166951. doi:10.1016/j.nima.2022.166951.
- Sotto-Maior Peralva, B. (2015). The TileCal Online Energy Estimation for the Next LHC Operation Period. *Journal of Physics: Conference Series*, 608, 012043. doi:10.1088/1742-6596/608/1/012043.
- Soyez, G. (2019). Pileup mitigation at the LHC: A theorist's view. *Physics Reports*, 803, 1–158. doi:10.1016/j.physrep.2019.01.007.
- Teixeira, T., Andrade, L., and de Seixas, J. (2021). Sparse deconvolution methods for online energy estimation in calorimeters operating in high luminosity conditions. *Journal of Instrumentation*, 16(09), P09008. doi:10.1088/1748-0221/16/09/P09008.
- Testa, M. and The ATLAS Collaboration (2019). Upgrade of the ATLAS detectors and trigger for the High Luminosity LHC: Tracking and timing for pile-up suppression. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 936, 394–395. doi:10.1016/j.nima.2018.10.172.
- The ATLAS Collaboration (2008). The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08), S08003–S08003. doi:10.1088/1748-0221/3/08/S08003.
- The ATLAS Collaboration (2010). Readiness of the ATLAS Tile Calorimeter for LHC collisions. *The European Physical Journal C*, 70(4), 1193–1236. doi:10.1140/epjc/s10052-010-1508-y.
- The ATLAS Collaboration (2020). Performance of the upgraded PreProcessor of the ATLAS Level-1 Calorimeter Trigger. *Journal of Instrumentation*, 3(08), S08003. doi:10.1088/1748-0221/15/11/P11016.
- Wigmans, R. (2017). *Calorimetry: Energy Measurement in Particle Physics*. Oxford University Press. doi:10.1093/oso/9780198786351.001.0001.
- Wong, T.T. and Yeh, P.Y. (2020). Reliable Accuracy Estimates from k -Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586–1594. doi:10.1109/TKDE.2019.2912815.