



Núcleo de
Computação
Científica



unesp

Relatório de Atividades

Thiago R. F. P. Tomei

— NCC/UNESP —

Relatório Anual
Período: Março/2023 a Março/2024

Índice

1	Resumo Executivo do Projeto de Pesquisa	2
2	Atividades de Pesquisa	3
2.1	Busca por Matéria Escura	3
2.2	Upgrade do Experimento CMS	9
2.3	Estudos em Aprendizado de Máquina	14
2.4	Outras Atividades	16
2.5	Bases de Dados	17
3	Atividades Docentes	18
3.1	Disciplinas Ministradas	18
3.2	Orientações Acadêmicas	18
3.3	Participação em Bancas	18
3.4	Organização de Escolas	18
4	Atividades de Extensão	19
4.1	MasterClass de Física de Altas Energias	19

1. Resumo Executivo do Projeto de Pesquisa

Para o triênio 2022–2024, planejo seguir me dedicando ao trabalho na colaboração *Compact Muon Solenoid* (CMS) [1], que opera o experimento de mesmo nome do *Large Hadron Collider* (LHC) [2]. O experimento CMS planeja tomar, durante o Run 3 do LHC, cerca de 150 fb^{-1} de dados de colisões próton-próton a uma energia de centro de massa $\sqrt{s} = 13.6 \text{ TeV}$. Em paralelo a essa atividade, a colaboração se prepara para a manutenção e os aprimoramentos do experimento para a chamada Fase 2, quando o LHC será atualizado para sua configuração de alta luminosidade, o *High-Luminosity Large Hadron Collider* (HL-LHC). Até o momento, todos os resultados produzidos pelo LHC confirmam de maneira cabal as previsões do modelo padrão da física de partículas [3]; o objetivo principal deste projeto é continuar analisando os dados da maneira mais minuciosa possível, em busca de indícios de novos fenômenos.

Meu trabalho será realizado em três frentes. Na frente da *Busca pela Matéria Escura*, trabalharei na análise dos supracitados dados do Run 2 em busca de evidências da produção de novas partículas, não previstas pela física do modelo padrão, que possam ser responsáveis pela chamada matéria escura. Na frente do *Upgrade do Experimento CMS*, trabalharei nos estudos e projetos para a Fase 2 do experimento, que entrará em operação a partir de 2029 e representará uma mudança radical no paradigma da Física de Altas Energias, em particular na área de tomada de dados. Finalmente, na frente de *Estudos em Aprendizado de Máquina*, continuarei minha pesquisa nesse assunto com vista a integrar as ferramentas e técnicas da área em meu trabalho.

2. Atividades de Pesquisa

Detalho a seguir minhas atividades em cada uma das frentes de trabalho do meu projeto de pesquisa.

2.1 Busca por Matéria Escura

O indício mais forte de Física Além do Modelo Padrão é a chamada matéria escura (*dark matter* – DM). Uma das principais evidências a favor da existência da DM é a observação de que, para vários tipos de objetos observáveis – estrelas, nuvens de gás, galáxias –, sua dinâmica é diferente daquela esperada apenas devido à atração gravitacional de outros objetos. Outra evidência para a existência da Matéria Escura é o ajuste global dos parâmetros do Modelo Padrão Cosmológico, que implica uma densidade de matéria não-bariônica $\Omega_c h^2 = 0.1196 \pm 0.0031$, *i.e.*, uma densidade física de Matéria Escura de cerca de 23% [4]. As características que definem um candidato a matéria Escura são a estabilidade em escalas de tempo cosmológicas, a interação extremamente fraca com a radiação eletromagnética e a densidade de relíquia correta. Nenhuma das partículas previstas pelo Modelo Padrão atende esses requisitos; várias extensões do Modelo Padrão, por sua vez, apresentam um ou mais candidatos a matéria escura. No LHC, o candidato a matéria escura χ poderia ser produzido em reações do tipo $pp \rightarrow \chi\chi + X$, onde a matéria escura deixa a região de interação sem deixar sinal no detector, aparecendo como momento transverso faltante (p_T^{miss}). Nesta frente de trabalho, me concentrei no estudo de modelos com partículas de vida longa (*long-lived particles* – LLP). A Figura 1 mostra um conjunto de possíveis assinaturas nessa categoria.

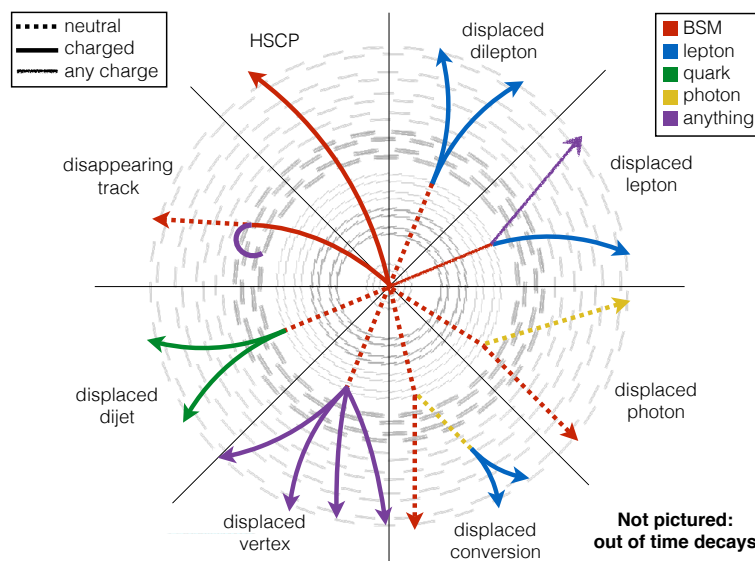


Figura 1: Diferentes assinaturas da produção de partículas de vida longa.

2.1.1 Modelos com Assinaturas de Traços Evanescentes

Nos modelos de nova física com assinaturas de *traços evanescentes*, um novo tipo de LLP carregada é produzida na colisão pp. Devido à essa longa vida média, a partícula percorre uma distância macroscópica antes de se desintegrar em partículas mais estáveis. Esses produtos de decaimento não são, entretanto, observados pelo detector; ou sua energia é baixa demais para sua reconstrução, ou eles interagem fracamente demais com a matéria usual, *i.e.*, são um candidato a matéria escura. Após a reconstrução, a assinatura final é de um traço que parece desaparecer antes do final do sistema de detecção.

Até agora, três buscas por traços evanescentes no CMS foram realizadas usando dados do Run 1 e Run 2 do LHC [5–7] e, com base nos resultados mais recentes, nenhuma evidência da existência de partículas de vida longa foi encontrada. Limites superiores para a massa/tempo de vida dos charginos (partícula carregada eletricamente do modelo de Quebra de Supersimetria Mediado por Anomalias - AMSB [8, 9] que vive tempo suficiente para deixar sinais nas camadas mais internas do tracker, denotado por $\tilde{\chi}_1^\pm$) foram obtidos, sendo 884 (474) GeV para um tempo de vida de 3 (0,2) ns, no cenário onde o neutralino (partícula eletricamente neutra do modelo AMSB que é candidata a matéria escura, denotada por $\tilde{\chi}_1^0$) é do tipo wino-like; e 750 (175) GeV para um tempo de vida de 3 (0,05) ns, no cenário onde o neutralino é do tipo higgsino-like.

No momento, estou orientando o Sr. Breno Orzari em seu doutorado, visando uma busca desse tipo com os dados do LHC que estão sendo tomados no Run 3. Essa busca é feita em colaboração com o grupo do Prof. Christopher Hill, da Ohio State University. O Sr. Orzari trabalhou em estudos de viabilidade para a simulação de eventos de sinal usando o software MadGraph5 (MG5) [10], com a adição de jatos extras ao processo. Dado os resultados positivos para a simulação, a produção centralizada de eventos de sinal, ou seja, realizada pelo grupo da colaboração CMS especializado nesse processo, foi solicitada e concluída. A Figura 2 mostra comparações entre eventos produzidos pelo estudante e aqueles produzidos centralmente para algumas variáveis de interesse da partícula de vida longa. É possível notar que ambos estão de acordo conforme esperado. A razão para a produção centralizada é devido ao alto volume de dados a ser gerado e para alcançar um alto grau de confiança, o que é estritamente necessário para continuar a análise de dados.

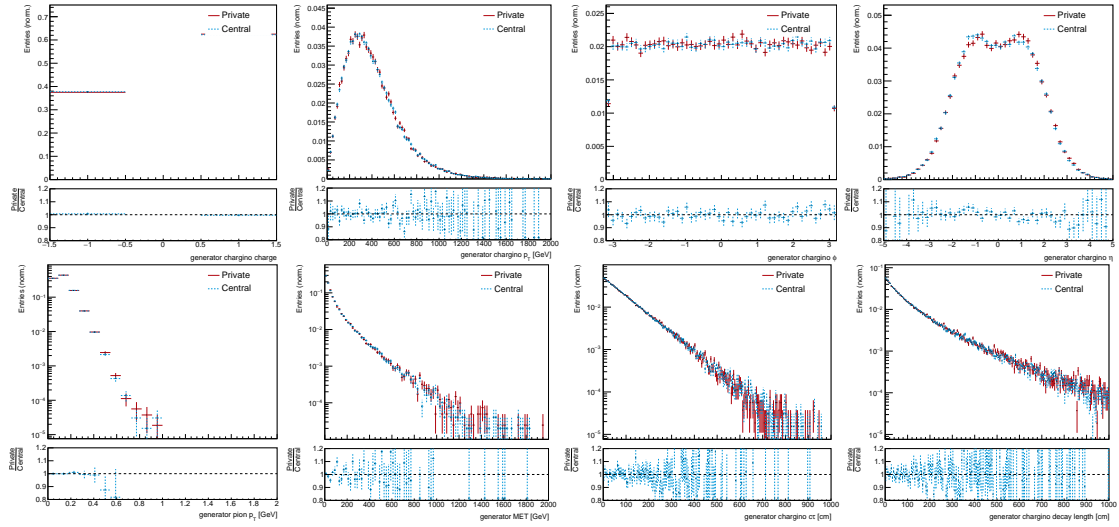


Figura 2: Comparação das características das partículas em amostras de sinal da geração privada (linha sólida vermelha) e geração centralizada (linha tracejada azul). Alto, da esquerda para a direita: carga elétrica, p_T , η and ϕ dos charginos gerados. Baixo, da esquerda para a direita: p_T dos píons do decaimento do $\tilde{\chi}_1^\pm$; p_T^{miss} , tempo de vida (em unidades de $c\tau$) e comprimento de decaimento dos charginos gerados.

O Sr. Orzari também realizou estudos sobre a eficiência do trigger para a análise. Como parte da seleção de eventos de interesse, um dos primeiros passos é verificar se o evento ativou ou não um determinado *caminho de trigger* (algoritmos para reconstruir e filtrar eventos com base em determinadas quantidades), sendo importante manter a boa qualidade desses caminhos durante a coleta de dados do CMS. A análise para a busca de traços evanescentes usa algoritmos baseados no p_T^{miss} medido pelos calorímetros, já que os traços que desaparecem não contribuem para o momento visível nesses detectores. A qualidade dos caminhos é medida por meio de sua eficiência, ou seja, a porcentagem de eventos filtrados dada uma variável de interesse.

A Figura 3 mostra a medição da eficiência de um algoritmo que seleciona eventos $p_T^{\text{miss}} > 105\text{GeV}$ e um traço isolado com $p_T > 50\text{GeV}$ (HLT_MET105_IsoTrk50). Na imagem à esquerda, a eficiência é calculada em função da energia transversa ausente excluindo múons usando eventos de múons coletados pelo CMS em 2022/2023 e eventos simulados de $W\beta\ell\nu$. Pode-se notar que o valor da eficiência não atinge 100%, dado que a própria eficiência de reconstrução de traços do algoritmo é em torno de 80%. Na imagem à direita, a eficiência é calculada em função do número de camadas medidas usando eventos de sinal produzidos centralmente pelo CMS para diferentes tempos de vida do chargino. Nota-se uma grande influência do comprimento do traço (diretamente ligado ao tempo de vida do chargino) na eficiência do caminho, como esperado.

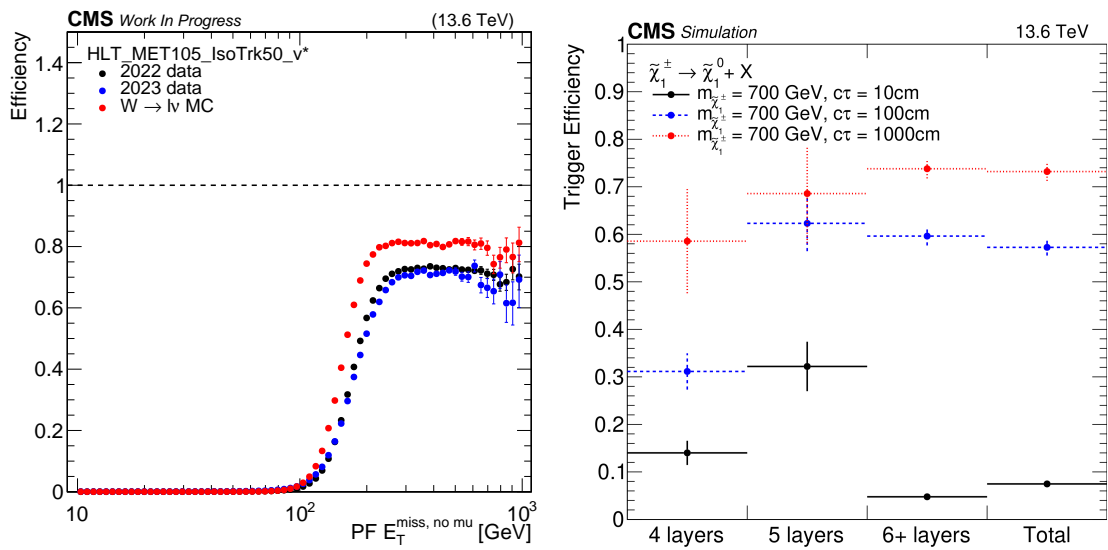


Figura 3: Cálculo de eficiência para o caminho HLT_MET105_IsoTrk50. Esquerda: em função da momento transverso faltante excluindo múons usando eventos de múons do CMS coletados em 2022/2023 (círculos pretos/azuis) e eventos simulados de $W \rightarrow l\nu$ (círculos vermelhos). Direita: em função do número de camadas com medições nos traços usando eventos simulados centralmente de charginos com massa de 700 GeV e 10 (linhas/círculos pretos sólidos), 100 (linhas/círculos azuis tracejados) e 1000 cm (linhas/círculos vermelhos pontilhados) de τ .

2.1.2 Modelos com Assinaturas de Vértices Deslocados

Uma outra assinatura possível de LLPs são os chamados vértices deslocados, assunto no qual estou orientando o Sr. Bruno Lopes em seu doutorado. Em particular, procuramos por vértices dentro do tubo do LHC onde os feixes são colididos (*beam pipe*). Nesse período, houve uma reinterpretação da análise da Run 2, a migração de um dos *triggers* de eventos utilizados e estão sendo feitos estudos preliminares para a Run 3.

A assinatura de vértices deslocados cobre uma variedade de modelos. A análise publicada pelo CMS com dados da Run 2 [11] busca por estados finais puramente hadrônicos. Em particular, é utilizado um *trigger* que seleciona eventos baseado na soma escalar da energia dos jatos hadrônicos detectados, H_T . Foi considerado um modelo supersimétrico com violação de R-paridade, testando dois processos: LLP neutra (neutralino), resultando em 4 jatos hadrônicos no estado final, ou LLP carregada (stop, parceiro supersimétrico do quark top), resultando em 6 jatos. Como o decaimento da LLP ocorre dentro do *beam pipe*, mesmo no caso em que a LLP é carregada, não há traço no detector e é observado somente o vértice deslocado.

Para a reinterpretação da análise da Run 2, foram utilizados dois tipos de modelo: um modelo simplificado com um mediador pesado, seja escalar ou vetorial, e uma classe de modelos supersimétricos denominada *stealth SUSY* [12]. No último caso, a LLP é um novo singleto que possui um parceiro supersimétrico ("singlino"), originado do decaimento de um stop e são analisadas as possibilidades de jatos originados de gluons ou de quarks b no estado final. Para a reinterpretação, foi usado exatamente o mesmo procedimento da análise da Run 2, com os mesmos critérios de seleção. Os resultados foram incluídos no artigo de revisão sobre análises que buscam matéria escura no CMS, em processo de publicação [13]. Em especial, a análise de vértices deslocados apresenta os maiores limites de exclusão para a massa do stop no modelo de *stealth SUSY* em que o estado final são jatos leves e o comprimento de decaimento é entre 0.1 mm e 1 mm.

Além do *trigger* baseado puramente em atividade hadrônica, outra estratégia é utilizar outro critério de seleção em paralelo, de maneira a permitir menores valores de H_T e ter sensibilidade a massas menores para a LLP. Para vértices deslocados, uma possibilidade é utilizar um algoritmo de classificação de jatos originados de decaimentos de mésons b , visto que seu comprimento de decaimento, principal parâmetro para a identificação, é parecido com o sinal desejado. O *trigger* em questão reduz o limite na variável H_T por um fator 3, ao preço de exigir 3 jatos classificados como originados de um méson b . Até 2023, a rede neural utilizada era a DeepJet, e foi feita a migração para a Particle Net [14], que apresenta melhor desempenho. Apesar disso, a sensibilidade ao sinal foi melhorada para massas menores, com a eficiência aumentando de 20% para 34% no modelo teste com violação de R-paridade e o stop como a LLP, com massa de 200 GeV e comprimento de decaimento de 2 mm.

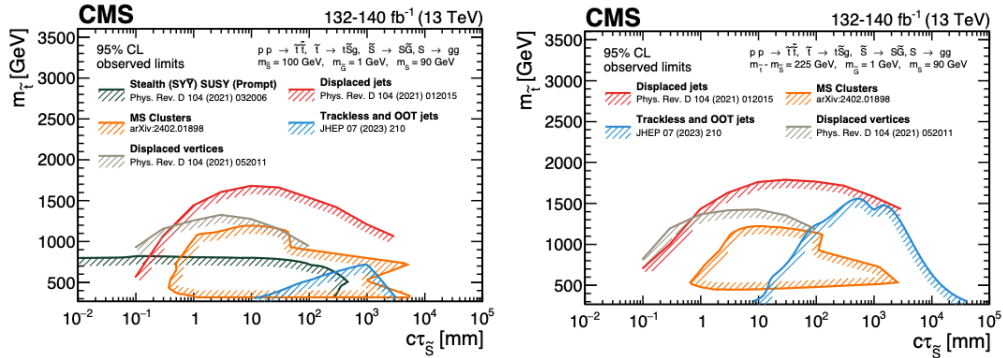


Figura 4: Limites de exclusão para o modelo de *stealth SUSY* de diferentes análises do CMS. Em cinza, o resultado para vértices deslocados. Na esquerda, o caso em que a massa do singlete é fixa em 100 GeV, e na esquerda, o caso em que a diferença de massa entre o stop e o singlete é fixa em 225 GeV.

Para a Run 3, é considerada a possibilidade do uso de *scouting* para reduzir o limite na variável H_T sem precisar de outro critério de seleção. *Scouting*, neste contexto, consiste em salvar menos informações dos eventos, de maneira que é possível salvar mais eventos. O *trigger* atualmente utilizado pelo CMS nesse sentido envolve vários critérios de seleção a nível de hardware e não há seleção a nível de software. Estudos preliminares utilizando o modelo com violação de R-paridade mostram eficiências próximas de 100% para massas da ordem de 200 GeV, e da ordem de 50% para massas da ordem de 100 GeV, uma ordem de grandeza acima do trigger que utiliza jatos b . No entanto, a análise da Run 2 utiliza um algoritmo customizado de identificação de vértices, algo que teria que ser modificado com a redução nas informações sobre os eventos, diminuindo a capacidade de distinguir o sinal dos eventos originados por processos do modelo padrão. Estudos posteriores serão dedicados a quantificar esta perda, assim como testar a viabilidade do uso de outras ferramentas, como uma rede neural, para distinguir tais eventos.

2.2 Upgrade do Experimento CMS

A partir de 2029, o LHC entrará em sua fase de alta luminosidade (HL-LHC), com luminosidade instantânea de até $7.5 \times 10^{34} / \text{cm}^2 / \text{s}$, condições essas que levarão a um *pileup*¹ médio $\langle \text{PU} \rangle = 200$ e uma luminosidade integrada final de 3000 fb^{-1} em 2038. O experimento CMS vai passar por uma série de aprimoramentos para fazer frente a essas novas condições na chamada *Fase 2* [15].

Um aspecto crítico da Fase 2 é o sistema de trigger e aquisição de dados [16]. Durante o período deste relatório, terminou meu mandato como **Coordenador do Upgrade do HLT para a Fase 2**. Desde setembro de 2022, passei a trabalhar como **Coordenador do Field Operations Group (FOG Convener)**, como mostrado na Figura ?? . Acredito que este é o momento certo para ocupar esta posição; como FOG Convener, sou o responsável pela coordenação da implantação e monitoramento do High-Level Trigger do experimento. Também lidero o coordeno o grupo de plantonistas (*High-Level Trigger Detector On-Call – HLT-DOC*) do sistema. Finalmente, ajo como segunda linha de suporte no caso de problemas com o trigger. Embora esta atividade não seja diretamente ligada ao *upgrade*, acredito que é importante que eu tenha esta experiência.

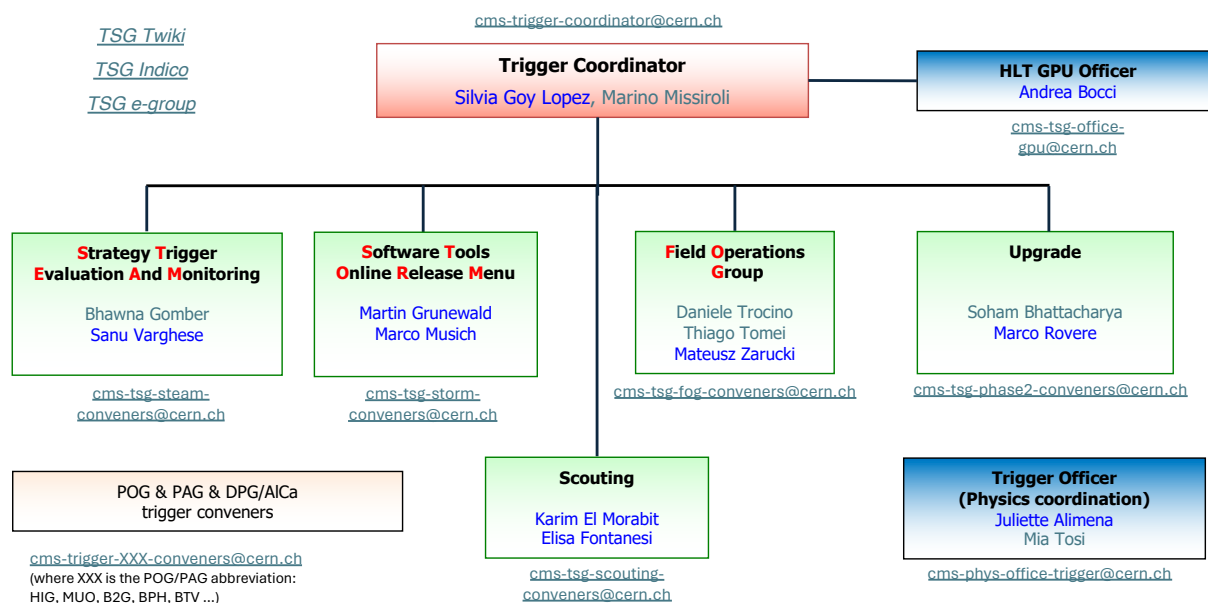


Figura 5: Organograma do Grupo de Estudos do Trigger, abril de 2024.

¹Ocorrência de múltiplas colisões nos cruzamentos de grupos de prótons do acelerador.

2.2.1 Next-Generation Triggers

Como parte de suas atividades na Atualização do Trigger de Alto Nível do CMS, eu me juntei ao projeto Next Generation Triggers (NGT)². O projeto visa aprimorar as capacidades dos experimentos CMS e ATLAS no LHC, melhorando os sistemas de trigger e aquisição de dados, e foca na integração de técnicas avançadas de computação, incluindo aprendizado de máquina e inteligência artificial de ponta, para otimizar a seleção e o processamento de eventos de colisão em tempo real. O projeto NGT foi aprovado pelo conselho do CERN em outubro de 2023 [17] e está organizado em quatro Pacotes de Trabalho (Working Packages – WP):

- *Infraestrutura, Algoritmos e Teoria (WP1)*, focado no desenvolvimento de ferramentas e metodologias de software para otimizar geradores de eventos, código dos experimentos e algoritmos de análise, incluindo aqueles baseados em redes neurais e em arquiteturas aceleradas.
- *Aprimoramento do Trigger e Aquisição de Dados do ATLAS (WP2) e Aprimoramento do Processamento de Dados em Tempo Real do CMS (WP3)*, pacotes de trabalho específicos para os experimentos que irão redesenhar suas estratégias de coleta de dados para expandir o alcance de seus programas de física.
- *Programas Educacionais e Divulgação (WP4)*, para promover intercâmbios entre cientistas da computação e pesquisadores de física, academia e indústria e complementar os programas educacionais existentes para treinar habilidades em ciência de dados e IA para a próxima geração de físicos de altas energias.

A Figura 6 mostra o organograma completo do NGT, até os quatro WPs. Eu estou atualmente atualmente envolvido no WP3, com o objetivo de ampliar o alcance do programa de física do CMS em vista da era do HL-LHC. Agora daremos uma breve visão geral desse Pacote de Trabalho e detalharemos sua participação nesses esforços.

WP3: Aprimoramento do Processamento de Dados em Tempo Real do CMS

Sabe-se que os experimentos do LHC podem processar apenas uma fração das colisões entregues, enquanto a maioria precisa ser descartada pelo sistema de seleção em tempo real (o sistema de trigger) devido a recursos de armazenamento limitados. O objetivo deste Pacote de Trabalho é remover essa limitação, aprimorando grandemente os sistemas de trigger do CMS, visando redesenhar a estratégia de coleta de dados para reduzir (e quando possível remover) a necessidade de rejeitar eventos e otimizar as informações armazenadas por evento. Pretendemos alcançar isso aproveitando algoritmos tradicionais baseados em física e soluções avançadas de IA, e duas direções serão

²<https://nextgentriggers.web.cern.ch/>

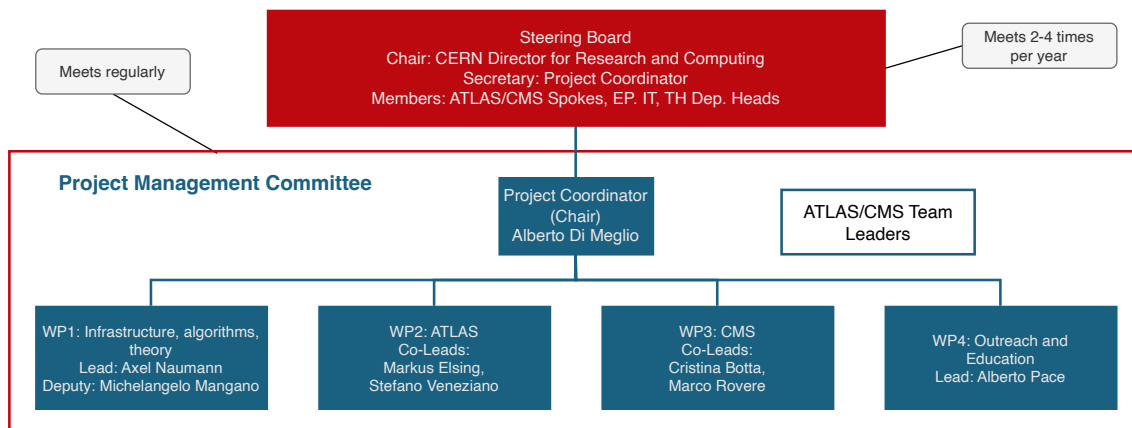


Figura 6: Organograma do Next Generation Triggers.

exploradas:

- Extensões e abordagens alternativas para o scouting com Level-1 Trigger (L1T), para aumentar a capacidade de realizar análises de física em todos os eventos de colisão apenas com informações do L1T.
- Um sistema de Trigger de Alto Nível de próxima geração (“Revolução na Reconstrução em Tempo Real” ou R³) capaz de realizar a reconstrução completa de eventos em tempo real em todos os eventos selecionados pelo L1T, sem filtração adicional antes do armazenamento dos dados.

Esses dois passos principais permitiriam à colaboração CMS remover o gargalo induzido pela seleção de eventos em tempo real e expandir nosso alcance tanto para descobertas e medições de precisão. Projetos de P&D e provas de conceito nessas atividades estão em andamento, e alguns deles já estabeleceram a validade das ideias por trás das tarefas propostas. Focamos aqui no subprojeto R³, que é subdividido em cinco tarefas³:

- 3.1) Desenvolver algoritmos modernos e estruturas de dados adaptadas a arquiteturas heterogêneas, para reconstrução de todos os objetos físicos do CMS.
- 3.2) Processamento distribuído para o HLT dentro do CMSSW, suportando o offloading de algoritmos para aceleradores remotos.
- 3.3) Investigar e desenvolver abordagens para redução do tamanho dos dados brutos, por exemplo, compressão de dados sem perda, com perda ou substituição de dados brutos por dados reconstruídos localmente. Ambas as abordagens já estão sendo investigadas para a tomada de dados de íons pesados.

³Usamos a mesma numeração de tarefas do projeto NGT.

- 3.4) Soluções para armazenar os dados de entrada do HLT por algumas horas, executar calibrações rápidas e ótimas online e usá-las para o processamento final do HLT.

Eu sou o líder da tarefa 3.4 – *Calibrações Ótimas*. O objetivo é projetar fluxos de trabalho de calibração acelerados para alcançar no HLT a mesma precisão que a reconstrução offline. Vamos otimizar o processo de calibração para os detectores do CMS introduzindo armazenamento de dados online e aproveitando técnicas de IA preditiva. Como tal, começamos a examinar todos os fluxos de trabalho de alinhamento e calibração no experimento. Esses incluem:

- A determinação da região luminosa das colisões, também conhecida como *beamspot* [18]. Seus parâmetros são determinados a partir de uma média de muitos eventos, em contraste com o vértice primário por evento que dá a posição precisa de uma única colisão, e são atualizados continuamente durante a tomada de dados. A posição do centro do beamspot é usada, especialmente no HLT, para estimar a posição do ponto de interação antes da reconstrução do vértice primário e para fornecer uma restrição adicional na reconstrução de todos os vértices primários de um evento.
- O Loop de Calibração Rápida (*Prompt Calibration Loop* – PCL) [19]. Durante a operação normal do experimento CMS no Run 2, cerca de 1 kHz de dados foram processados no processamento em massa. Nesse cenário, apenas uma largura de banda limitada, correspondente a cerca de 10% do fluxo de dados, é alocada para um processamento expresso a fim de garantir uma reconstrução rápida. Uma seleção de dados dos fluxos expresso e de calibração é usada para calcular as constantes de calibração (condições) atualizadas para um determinado run enquanto a maior parte dos dados é armazenada em disco. Dessa forma, a reconstrução inicial dos dados pode se beneficiar das melhores condições, reduzindo a necessidade de reprocessamento offline dos dados.
- A calibração rápida do calorímetro eletromagnético do CMS (ECAL), necessária para mitigar os efeitos das mudanças induzidas pelo envelhecimento na resposta dos canais, foi recentemente atualizada para um maior grau de automação [20]. O novo sistema, desenvolvido durante a segunda parada longa do LHC, é baseado em ferramentas padrão da indústria (OpenShift, Jenkins, Influxdb e Grafana) e fornece uma interface geral para orquestrar fluxos de trabalho autônomos escritos em diferentes linguagens de programação. Também fornece interfaces para outros sistemas existentes do CMS para conduzir o processamento de fluxos de dados selecionados e carregar

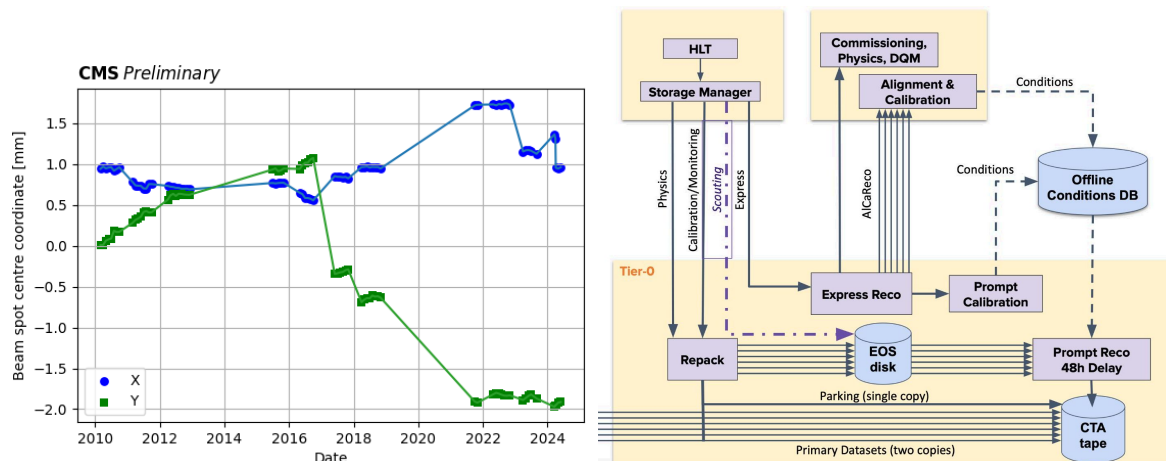


Figura 7: Esquerda: a evolução das coordenadas transversais do centro do beamspot do CMS em função do tempo. A reconstrução do beamspot é feita quase em tempo real, para atender às necessidades das reconstruções HLT e offline. Direita: o Loop de Calibração Rápida, embutido na maioria dos fluxos de trabalho de alinhamento e calibração no experimento CMS.

novas calibrações calculadas no banco de dados usado para o processamento de dados para análises de física.

A Figura 7 mostra a evolução do beamspot do CMS e o Fluxo de Trabalho de Calibração Rápida. Um levantamento completo dos fluxos de trabalho de alinhamento e calibração no experimento está em andamento.

Outro exercício que precisa ser feito para o projeto NGT é a estimativa dos requisitos potenciais em termos de recursos computacionais. Considerando um cenário de implantação para o Run 4, fazemos as seguintes suposições, com base nas estimativas do TDR do DAQ e HLT da Fase 2 do CMS [16]:

- A necessidade de um buffer de dados de 8 horas. Esse requisito surge da existência de SSDs de grau empresarial classificadas para “três gravações por unidade por dia” (DWPD) para um determinado número de anos, e a necessidade de operação contínua – $24 \text{ horas} / 3 = 8 \text{ horas}$.
- Uma taxa de L1T de 750,kHz. Embora o cenário do Run 4 preveja apenas uma taxa de L1T de 500,kHz, a experiência mostra que o CMS tradicionalmente ultrapassa as especificações de design. Portanto, consideramos mais seguro adotar o cenário do Run 5.
- Consideramos um tamanho de evento de 6,1 MB, de acordo com as estimativas do TDR.
- Consideramos um fator de segurança de 50%.

Essas suposições levam a $8 \text{ horas} \times 750 \text{ kHz} \times 6,1 \text{ MB/evento} \times 1,5 = 200 \text{ PB}$. Considere-

ramos que o próprio cluster de computadores do HLT estará executando a reconstrução do NGT, que de acordo com o TDR são 1600 servidores montados em rack 1U; usaremos uma estimativa mais conservadora de 400 servidores 4U. Dessa forma, cada nó precisará hospedar $200 \text{ PB} / 400 = 500 \text{ TB}$. Considerando que SSDs de 25 TB, 3 DWPD já estão disponíveis no mercado, os requisitos de armazenamento podem ser atendidos com 20 discos por servidor.

Notamos que essa tarefa tem sinergia com as operações do Run-3, então estamos comprometidos em implantar um demonstrador aplicado ao scouting do HLT durante o último ano do Run 3. Um demonstrador NGT mínimo receberia uma cópia da entrada do scouting do HLT, reduzida por um fator $\mathcal{O}(100)$ – de $\sim 30 \text{ kHz}$ para $\sim 300 \text{ Hz}$. Esses dados seriam enviados para uma área especial no site experimental do CMS (P5), onde seriam armazenados por várias horas. Em seguida, usaríamos os fluxos de trabalho regulares para derivar um subconjunto das calibrações necessárias, seja no P5 dentro de 8 horas se nosso trabalho na tarefa 3.4 convergir a tempo, ou usando o PCL dentro de 48 horas se não convergir – mas observem-se os requisitos adicionais de armazenamento necessários para isso. Em seguida, reexecutaríamos o scouting do HLT nos dados armazenados com essas calibrações atualizadas, e compararíamos o desempenho da reconstrução original com a atualizada.

Estamos atualmente no processo de formação do grupo de trabalho. O Dr. Marco Musich (INFN–Pisa) juntou-se como especialista em alinhamento e calibração do detector para supervisionar o trabalho, e a Sra. Jessica Prendi (ETH–Zurich) juntou-se como estudante de doutorado para fornecer a força de trabalho, executando fluxos de trabalho de calibração e comparando seus resultados. Do lado da Unesp, também queremos contratar um bolsista de Treinamento Técnico e um estudante de doutorado; o candidato para esta última posição é o Sr. Cauê Evangelista de Sousa, sob a bolsa FAPESP 2024/10820-0.

2.3 Estudos em Aprendizado de Máquina

Em 2017, comecei a me dedicar a estudos em aprendizado de máquina (*machine learning* – ML), haja vista a aceleração do uso dessas técnicas em praticamente todas as áreas da ciência, e particularmente em física de altas energias. A partir de 2018 decidi estudar os chamados *modelos generativos*, tendo como objetivo foi a investigação de métodos de aceleração para a simulação de eventos de colisão de partículas, uma tarefa fundamental para o trabalho da comunidade de HEP. A simulação dos efeitos do detector em eventos de HEP geralmente é feita com o software GEANT4 [21], e pode levar mais de um minuto por evento para interações complexas como produção de $t\bar{t}$ no LHC na presença de pileup. Trabalhos anteriores nessa área

resultaram em abordagens como o ATLAS FastCaloSim [22] e o CMS FastSim [23], no qual os autores relatam grandes acelerações na simulação do calorímetro, mas à custa de algum sacrifício na precisão.

Os modelos geradores, por outro lado, são uma classe de modelos de ML que podem alcançar altas velocidades, mantendo alta precisão. Juntamente com a equipe de computação do SPRACE – Dr. Raphael Cobe e Sr. Jefferson Fialho – fiz uma investigação preliminar sobre dois tipos de modelos generativos: Redes Generativas Adversariais (*Generative Adversarial Networks – GANs*) e AutoCodificadores Variacionais (*Variational AutoEncoders – VAEs*). Além disso, também envolvi meu estudante Sr. Breno Orzari, que em uma visita ao CERN no final de 2019 estabeleceu uma colaboração com o Dr. Maurizio Pierini, líder do grupo *Machine Learning for Particle Physics (MPP-HEP)*. Nessa colaboração buscamos gerar jatos hadrônicos, uma das assinaturas mais comuns de eventos do LHC. O artigo citado no relatório anterior foi finalizado e publicado [24]. Alguns de nossos resultados estão mostrados na Figura 8.

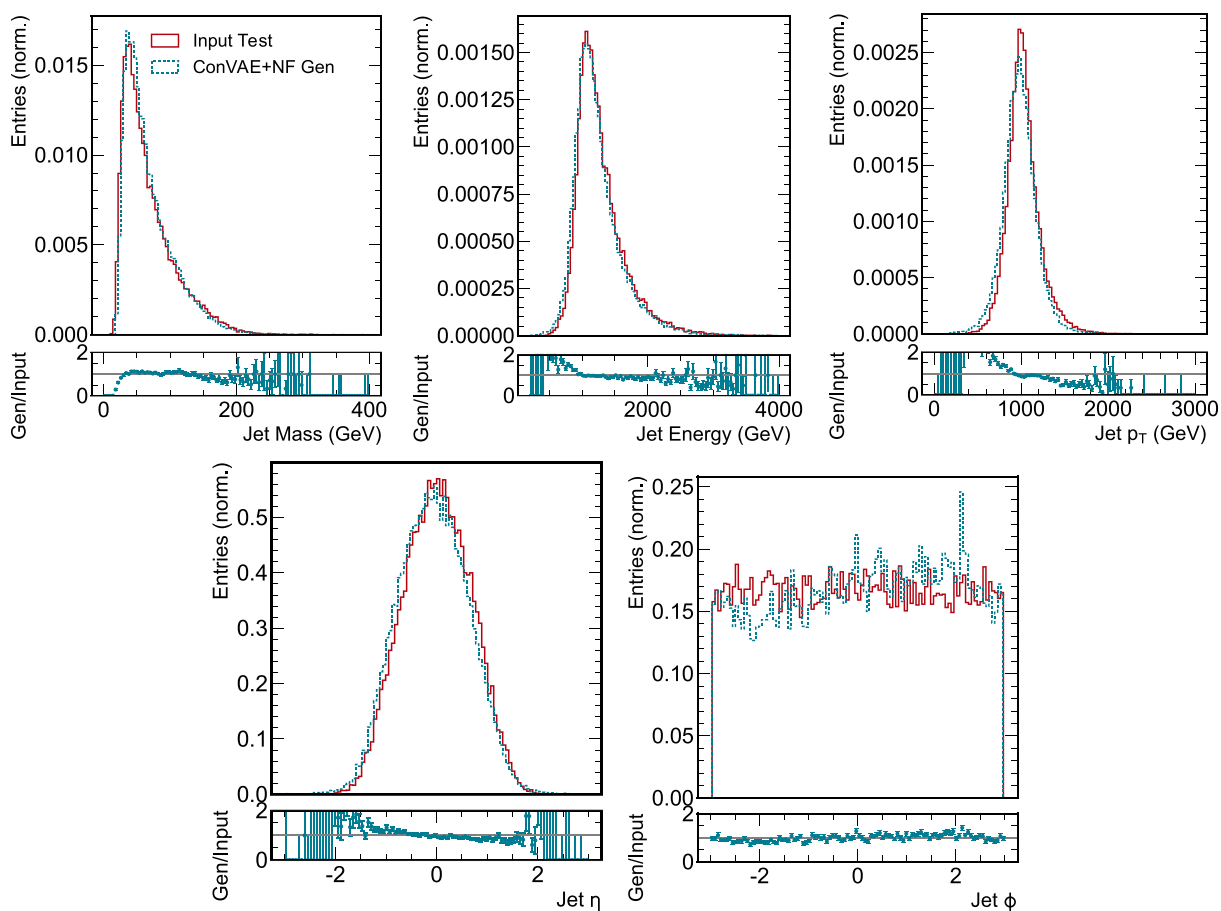


Figura 8: Resultados finais para geração de jatos com autocodificadores variacionais, para diversas variáveis. Alto, esquerda: energia dos jatos. Alto, direita: massa dos jatos. Baixo, esquerda: momento transversal dos jatos. Baixo, direita: pseudorapidez dos jatos.

2.4 Outras Atividades

2.4.1 CMS Collaboration Service Work

Como membro da colaboração CMS, é parte do meu trabalho científico contribuir para o bom andamento do experimento. Em 2023 participei das seguintes atividades de serviço para a colaboração:

- **Coordenador do Field Operations Group:** Desde setembro de 2022, em conjunto com o Drs. Mateusz Zarucki (Notre Dame) e Daniele Trocino (INFN-Pisa), coordeno o grupo do Operações de Campo do Trigger, tendo um papel fundamental na tomada de dados.
- **Líder da Tarefa de Calibrações Ótimas do Projeto “Next-Generation Triggers”:** Desde janeiro de 2024, lidero a respectiva tarefa, descrita na seção anterior, coordenando um pesquisador e um estudante no CERN pagos com recursos do projeto NGT. Também estou integrando essa frente de trabalho com meu projeto FAPESP no Brasil, e planejo contratar a mesma quantidade de pessoas trabalhando com bolsas brasileiras.

Minha contribuição também inclui participação direta na tomada de dados como HLT-DOC. Em 2023, dei plantões como HLT-DOC nas seguintes datas: 10/julho a 24/julho, e 21/agosto a 28/agosto, totalizando 21 dias. Também participei como Shift Leader, coordenando a operação do CMS diretamente do sítio experimental (P5). Em 2023, dei plantões como Shift Leader nas seguintes datas: 27/maio a 07/junho, totalizando 9 dias.

2.4.2 Participação em Conferências

- “The CMS High-Level Trigger for LHC Run-3”, apresentação oral no encontro *EPSBF 2023: III Encontro de Primavera da Sociedade Brasileira de Física*, Niteroi, Rio de Janeiro, Brasil (2022).

2.4.3 Fontes de Financiamento

Sigo sendo outorgado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) com um Auxílio à Pesquisa - Projeto Inicial (PI), do Programa Especial PNGP - Programa Nova Geração de Pesquisadores, intitulado *Computing Challenges for the CMS Phase-II Upgrade*, sob número de processo 2022/02950-5.

Sigo sendo outorgado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com uma bolsa de Produtividade em Pesquisa – PQ2, *Busca por Nova Física no Experimento CMS do Large Hadron Collider*, sob números de processo 314608/2020-6 e 307048/2023-3.

2.4.4 Artigos Completos Publicados em Periódicos

- **Artigos publicados com a Colaboração CMS:**

86 artigos publicados em 2023.⁴

Infelizmente, devido à inadequação do sistema eletrônico do CNPq, a lista de artigos da colaboração CMS não pode ser publicada no Lattes, pois é muito extensa.

- R. Kansal, A. Li, J. Duarte, N. Chernyavskaya, M. Pierini, B. Orzari, T. Tomei, *Evaluating generative models in high energy physics*, [Phys. Rev. D **107**, 076017 \(2023\)](#)
- B. Orzari, N. Chernyavskaya, R. Cobe, J. Duarte, J. Fialho, D. Gunopulos, R. Kansal, M. Pierini, T. Tomei, M. Touranakou, *LHC Hadronic Jet Generation Using Convolutional Variational Autoencoders with Normalizing Flows*, [Mach. Learn.: Sci. Technol. **4** 045023 \(2023\)](#)

2.5 Bases de Dados

- **Currículo Lattes:**

<http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4713328U4>

- **ORCID:**

<https://orcid.org/0000-0002-1809-5226>

- **Google Scholar:**

https://scholar.google.com.br/citations?user=Ud_lv6EAAAAJ&hl=en

Adicionalmente, os artigos completos publicados em 2023 estão listados na base de dados INSPIRE, acessível através do seguinte link: <https://inspirehep.net/literature?sort=mostrecent&size=100&page=1&q=find%20a%20tomei%20and%20cn%20cms%20and%20jy%202023%20and%20de%203E%202022-01-01>

⁴Cabe lembrar que a Colaboração CMS é formada por institutos de diversos países, entre os quais Rússia, Belarus e Ucrânia. A invasão da Ucrânia pela Rússia em 2022 levantou questões de ordem legal sobre as publicações da colaboração; em vista disso, as publicações foram temporariamente suspensas até que essas questões fossem adequadamente tratadas.

3. Atividades Docentes

3.1 Disciplinas Ministradas

- *Mecânica Quântica Avançada*: Instituto de Física Teórica (IFT-UNESP), primeiro semestre de 2024.

3.2 Orientações Acadêmicas

- *Sr. Breno Orzari*, estudante de Doutorado (2020–presente). Discente do Instituto de Física Teórica (IFT-UNESP).
- *Sr. Bruno Lopes*, estudante de Doutorado (2022–presente). Discente do Instituto de Física Teórica (IFT-UNESP).
- *Srta. Julia Carvalho Leite*, estudante de Doutorado (2022–presente). Discente do Instituto de Física Teórica (IFT-UNESP) (Coorientador).
- *Sr. Caio Evangelista de Sousa*, estudante de Mestrado (2022–2024). Discente do Centro de Ciências Naturais e Humanas (CCNH-UFABC) (Coorientador).

3.3 Participação em Bancas

- Marcio de Sousa Mateus Junior. Tese de Doutorado: *Produção de matéria escura em ressonância através de um novo mediador massivo de spin-1*, 2024. (Física). Universidade Federal do Rio Grande do Sul. Orientador: Gustavo Gil da Silveira.
- Matheus Costa Reis. Tese de Doutorado: *Direct production of $J/\psi\phi$ vector mesons at the LHCb experiment without additional activity*, 2023. (Física). Universidade Federal do Rio de Janeiro. Orientador: Murilo Santana Rangel.
- Carlos Yosep Bautista Choque. Tese de Doutorado: *Exploring Composite Higgs Models at Present and Future Colliders: A study on the Top-Higgs Sector and the Importance of Three-Body Decays in Vector-like Quark Searches*, 2023 (Física). Universidade Estadual Paulista. Orientador: Ricardo D'Elia Matheus.

3.4 Organização de Escolas

- *Second School on Data Science and Machine Learning*, escola organizada pelo IFT-Unesp, ICTP-SAIFR e AI2 de 04 a 08 de dezembro de 2023. <https://www.ictp-saifr.org/dsml2023/>.

4. Atividades de Extensão

4.1 MasterClass de Física de Altas Energias

Todos os anos, participo do *International MasterClass de Física de Altas Energias* [25], com participação de centenas de estudantes de ensino médio. Apresento o seminário *A Estrutura Elementar da Matéria*, que dá aos estudantes uma visão geral sobre o tema, e auxilio os estudantes na atividade prática do evento – a caracterização de eventos reais tomados pelo detector CMS em termos de sua topologia e sua classificação como eventos advindos de bósons W , Z ou Higgs. No período deste relatório, participei da edição de 2024 no Instituto de Física Teórica.

Gostaria de fazer aqui um aparte sobre o MasterClass Feminino. O dia 11 de fevereiro foi instituído como o Dia Internacional das Mulheres e Meninas na Ciência pela resolução A/RES/70/212, da Assembléia Geral das Nações Unidas [26]. O objetivo desse ato é chamar a atenção para a desigualdade de gênero e aumentar a participação feminina na área de ciências. De acordo com estudos patrocinados pela ONU, a probabilidade de uma estudante do sexo feminino obter um título de Bacharel, Mestre ou Doutora na área de ciências é de 18%, 8% e 2% respectivamente; as probabilidades correspondentes para estudantes do sexo masculino são de 37%, 18% e 6%. Tendo em vista a situação supracitada, o grupo SPRACE tem tomado a iniciativa de realizar essa edição especial do evento destinada especialmente a alunas de ensino médio e primeiro ano de licenciatura.

Finalmente, gostaria de também citar que estendi minhas atividades do Master-Class para meus colaboradores de fora do estado, proferindo a palestra “Os Aceleradores e as Experiências do CERN” no 1º Masterclass Física de Partículas da Universidade Tecnológica Federal do Paraná - Câmpus Toledo, realizado no período de 30/10/2023 a 31/10/2023.

Referências

- [1] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, [doi:10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [2] e. Evans, Lyndon and e. Bryant, Philip, “LHC Machine”, *JINST* **3** (2008) S08001, [doi:10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [3] Particle Data Group Collaboration, “Review of Particle Physics”, *PTEP* **2022** (2022) 083C01, [doi:10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [4] Planck Collaboration, “Planck 2018 results. VI. Cosmological parameters”, *Astron. Astrophys.* **641** (2020) A6, [doi:10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910), [arXiv:1807.06209](https://arxiv.org/abs/1807.06209). [Erratum: *Astron. Astrophys.* 652, C4 (2021)].
- [5] CMS Collaboration, “Search for disappearing tracks in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **01** (2015) 096, [doi:10.1007/JHEP01\(2015\)096](https://doi.org/10.1007/JHEP01(2015)096), [arXiv:1411.6006](https://arxiv.org/abs/1411.6006).
- [6] CMS Collaboration, “Search for disappearing tracks as a signature of new long-lived particles in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **08** (2018) 016, [doi:10.1007/JHEP08\(2018\)016](https://doi.org/10.1007/JHEP08(2018)016), [arXiv:1804.07321](https://arxiv.org/abs/1804.07321).
- [7] CMS Collaboration, “Search for disappearing tracks in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **806** (2020) 135502, [doi:10.1016/j.physletb.2020.135502](https://doi.org/10.1016/j.physletb.2020.135502), [arXiv:2004.05153](https://arxiv.org/abs/2004.05153).
- [8] G. F. Giudice, M. A. Luty, H. Murayama et al., “Gaugino mass without singlets”, *JHEP* **12** (1998) 027, [doi:10.1088/1126-6708/1998/12/027](https://doi.org/10.1088/1126-6708/1998/12/027), [arXiv:hep-ph/9810442](https://arxiv.org/abs/hep-ph/9810442).
- [9] L. Randall and R. Sundrum, “Out of this world supersymmetry breaking”, *Nucl. Phys. B* **557** (1999) 79–118, [doi:10.1016/S0550-3213\(99\)00359-4](https://doi.org/10.1016/S0550-3213(99)00359-4), [arXiv:hep-th/9810155](https://arxiv.org/abs/hep-th/9810155).
- [10] J. Alwall, R. Frederix, S. Frixione et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, [doi:10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079), [arXiv:1405.0301](https://arxiv.org/abs/1405.0301).
- [11] CMS Collaboration, “Search for long-lived particles decaying to jets with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **104** (2021), no. 5, 052011, [doi:10.1103/PhysRevD.104.052011](https://doi.org/10.1103/PhysRevD.104.052011), [arXiv:2104.13474](https://arxiv.org/abs/2104.13474).
- [12] J. Fan, M. Reece, and J. T. Ruderman, “Stealth Supersymmetry”, *JHEP* **11** (2011) 012, [doi:10.1007/JHEP11\(2011\)012](https://doi.org/10.1007/JHEP11(2011)012), [arXiv:1105.5135](https://arxiv.org/abs/1105.5135).
- [13] CMS Collaboration, “Dark sector searches with the CMS experiment”, [arXiv:2405.13778](https://arxiv.org/abs/2405.13778).

- [14] H. Qu and L. Gouskos, “ParticleNet: Jet Tagging via Particle Clouds”, *Phys. Rev. D* **101** (2020), no. 5, 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.
- [15] CMS Collaboration, “Technical Proposal for the Phase-II Upgrade of the CMS Detector”. <https://cds.cern.ch/record/2020886>, 2015.
- [16] CMS Collaboration, “The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger”. <https://cds.cern.ch/record/2759072>, 2021.
- [17] “TWO-HUNDRED-AND-THIRTEENTH SESSION OF THE COUNCIL”. <https://council.web.cern.ch/en/content/213>.
- [18] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [19] CMS Collaboration, F. De Guio, “CMS Alignment and Calibration workflows: lesson learned and future plans”. <https://cds.cern.ch/record/2121268>, 2016.
- [20] CMS Collaboration, S. Pigazzini, “Automatic data processing for prompt calibration of the CMS ECAL”. <https://cds.cern.ch/record/2853679>, 2023.
- [21] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth.* **A506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [22] ATLAS Collaboration, “The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim”. <https://cds.cern.ch/record/1300517>, 2010.
- [23] CMS Collaboration, “The fast simulation of the CMS detector at LHC”, *J. Phys. Conf. Ser.* **331** (2011) 032049, doi:10.1088/1742-6596/331/3/032049.
- [24] B. Orzari, N. Chernyavskaya, R. Cobe et al., “LHC hadronic jet generation using convolutional variational autoencoders with normalizing flows”, *Mach. Learn. Sci. Tech.* **4** (2023) 045023, doi:10.1088/2632-2153/ad04ea, arXiv:2310.13138.
- [25] “International Masterclasses – Hands on Particle Physics”. <http://www.physicsmasterclasses.org/>.
- [26] U. Nations, “International Day of Women and Girls in Science”. <http://www.un.org/en/events/women-and-girls-in-science-day/>.