

Um Classificador Neural *On-line* para a Física de Partículas

Denis O. Damazio¹, Paulo V. M. Silva¹, José M. Seixas¹

¹Laboratório de Processamento de Sinais (LPS)

COPPE/EE/UFRJ, CP 68504, Rio de Janeiro 21945-970, Brazil

E-mails: damazio@lps.ufrj.br, vitor@lps.ufrj.br, seixas@lps.ufrj.br

Abstract

An on-line neural system is being developed for detecting outsiders in experimental high-energy particle beams. This detection is based on exploring the information provided by a calorimeter, which is one of the main detectors used in modern particle collider experiments, as it measures the energy of the incoming particles. The methodology for training the network with incoming information is described and the discrimination of outsiders pions, muons and electrons are achieved with an efficiency better than 93%, from a simulation of the on-line system. The high quality of the neural detection is attested from a classical off-line analysis often used by experts in calorimetry. Implementation of the system on digital signal processor technology is also evaluated.

1 Introdução

De forma a explorar as propriedades mais fundamentais da matéria e estudar os fenômenos que participam da sua formação, cientistas do mundo inteiro têm que estudar e utilizar técnicas e instrumentos modernos. No CERN, Centro Europeu de Pesquisa Nuclear [1], localizado em Genebra, Suíça, estão sendo utilizados poderosos aceleradores de partícula para criar, por pequenos instantes de tempo, uma imensa concentração de energia num espaço o mais reduzido possível. A idéia é que estas seriam as mesmas condições presentes no início da formação do universo, onde várias forças e partículas diferentes estavam mais evidentes. Com a expansão e conseqüente resfriamento do universo, estas mesmas partículas tornam-se mais difíceis de serem detectadas, ao formarem outras partículas, mascarando também fatos básicos que constituem a matéria.

Assim, para que se possa pesquisar mais profundamente a estrutura da matéria, vários centros na Europa e nos Estados Unidos disputam a produção de aceleradores cada vez mais poderosos. Atualmente, o CERN está encerrando as operações de seu gigantesco acelerador de partículas, o LEP (Grande Colisionador Elétron-Pósitron), para iniciar a montagem do LHC (Grande Colisionador de Hadrons), que deve entrar em operação em 2005. Este novo acelerador gerará dois feixes de partícula a velocidades enormes e sentidos contrários, colidindo estes feixes em pontos específicos. Espera-se que, graças aos vários anéis magnéticos supercondutores, a enorme

energia de 14 TeV seja atingida no centro de gravidade.

Para ter acesso ao que ocorre no ponto de colisão, deve-se estudar os subprodutos originados. Assim, em volta dos dois pontos de colisão já previstos para o LHC, serão dispostos dois conjuntos de detectores com vários subdetectores especializados em cada tipo de partícula que possa surgir. Um destes detectores é o ATLAS, que pode ser visto na Figura 1.

O ATLAS será disposto como um barril, detectando as partículas que virão da linha do feixe (eixo central do barril). Um dos detectores do ATLAS é o calorímetro hadrônico, que mede a energia das partículas que nele incidem, por absorção das mesmas. Este calorímetro, que também formará um barril em torno do ponto de colisão, é dividido em 64 módulos em formato de cunha, que são desenvolvidos e testados separadamente.

Durante as fases de teste, feixes de diferentes partículas são feitos colidir contra o módulo do calorímetro. A partir destes testes, são feitas variadas calibrações e desenvolvimentos do sistema associado ao calorímetro. Entretanto, há um problema crucial envolvendo estes testes. O mesmo processo utilizado para gerar determinados tipos de partícula gera também outros tipos de partículas. Como algumas calibrações dependem que se tenha apenas um tipo de partícula para se analisar, variadas técnicas são utilizadas para se retirar do conjunto de dados tais eventos considerados indesejados. O uso de redes neurais tem sido proposto [2] para se identificar quais são os eventos desejados, demonstrando uma metodologia (de análise de eventos *off-line*) eficaz na eliminação desta contaminação. O presente trabalho é um estudo da aplicação desta metodologia de forma *on-line*, ou seja, a rede neural validaria os eventos enquanto estes ainda estivessem trafegando no sistema de aquisição de dados. Tal aplicação resultaria em um uso mais eficaz do custoso feixe de partículas, pois o sistema de aquisição de dados só armazenaria os eventos que fossem realmente de interesse para o experimento.

Para que possamos melhor entender o sistema desenvolvido, discutiremos na próxima seção alguns detalhes do calorímetro. Na terceira seção discutiremos a metodologia para o sistema *on-line* e também apresentaremos os resultados já obtidos na escolha do melhor conjunto de parâmetros necessários ao rápido treino da rede. Na quarta seção discutiremos os resultados obtidos com uma simulação da rede, em um Processador Digital de Sinais (DSP) e em um PC, visando acelerar o processamento.

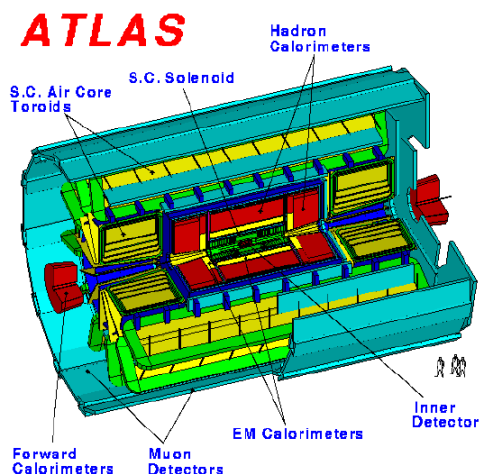


Figura 1: Conjunto de detectores do ATLAS

2 O Calorímetro

Desde 1996, vêm sendo testados protótipos de última geração para o calorímetro hadrônico do ATLAS. Os feixes que têm sido utilizados nos testes experimentais deste calorímetro são de três tipos diferentes de partículas: elétrons, píons e múons. Entretanto, como já havíamos apontado anteriormente, estes feixes contêm um certo grau de contaminação, havendo tipicamente, múons no feixe de píons e, píons e múons no feixe de elétrons.

Este calorímetro é formado basicamente por dois materiais. Um material pesado, no caso ferro, para servir de barreira às partículas que estão entrando, e um material ativo, telhas de material cintilante, que transforma a energia das reações nucleares ocorridas em luz, permitindo uma leitura da mesma. A partícula do feixe, ao entrar no material pesado, colide contra os átomos deste material, vindo a perder parte de sua energia e a se transformar em outras partículas de menor energia. Estas últimas colidem com outros átomos do material pesado e o processo se repete, formando uma cascata de partículas que se distribui na estrutura do calorímetro, que, então absorve inteiramente a energia da partícula incidente.

A topologia desta cascata de partículas mantém uma forte correspondência com a partícula inicial, podendo ser usada para identificar esta última [3]. Fibras óticas carregam a luz gerada para fora do módulo. Essas fibras são organizadas de forma que um grupo de fibras, relacionadas a uma dada região do calorímetro, seja unida, formando células de leitura na superfície do calorímetro. A Figura 2 traz a organização destas células, que chegam a 23 para cada metade de cada um dos 64 módulos do calorímetro. Como cada um dos dois lados da mesma célula são lidos separadamente, temos um total de 46 informações sendo lidas simultaneamente, para cada evento. Visando uma operação *on-line* do discriminador neural que está sendo proposto, as leituras duplas de cada célula foram compostas para gerar a média geométrica de energia correspondente a cada célula. Desta maneira, o

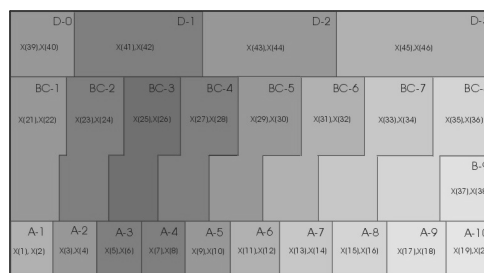


Figura 2: Segmentação do calorímetro. Vide texto.

discriminador pode ser projetado com vetores de apenas 23 componentes, se tornando mais compacto e, assim, mais veloz.

3 Metodologia Estudada

As 23 informações geradas por cada evento são utilizadas na entrada de uma rede neural, a ser treinada com o método *backpropagation* [4], utilizando como função de ativação, a tangente hiperbólica. Antes de alimentar a rede, estes valores de energia são normalizados pela sua soma, de forma a se estudar o processamento neural o mais independentemente possível da energia do feixe, concentrando todo o esforço de análise na topologia do problema.

Para a operação *on-line*, a rede neural deverá ser treinada dinamicamente, sendo alimentada pela informação de energia de partículas do feixe e correspondente contaminação. Por exemplo, para feixes de elétrons, píons e múons de contaminação dificultam o aprendizado requerido. No caso do feixe de píons, múons de contaminação estarão distorcendo o aprendizado, pois estarão sendo etiquetados como múons na fase de aprendizado. Como o processamento neural foi capaz de detectar tais partículas de contaminação na análise *off-line*, a partir do perfil de deposição de energia para cada partícula, espera-se que o processamento neural também tenha sucesso no caso *on-line* mais complexo.

Como forma de enfrentar tal complexidade da tarefa de discriminação *on-line*, particiona-se tal tarefa, identificando-se, primeiramente os múons de contaminação no feixe. A metodologia desenvolvida para o processamento *on-line* pode ser, então, descrita na seguinte forma :

- Inicialmente, são adquiridos os múons, a partir de feixe próprio. Estas partículas são consideradas livres de qualquer contaminação significativa.
- Os píons começam a ser adquiridos. Ao mesmo tempo (*on-line*), inicia-se o treinamento de uma rede neural com apenas uma saída, sendo treinada para que píons atinjam nesta saída o valor 1 e os múons -1. Os múons de contaminação no feixe de píons são assim identificados, como veremos adiante.
- Após o treinamento com os píons, pode-se retirar deste conjunto a contaminação por múons. Começa

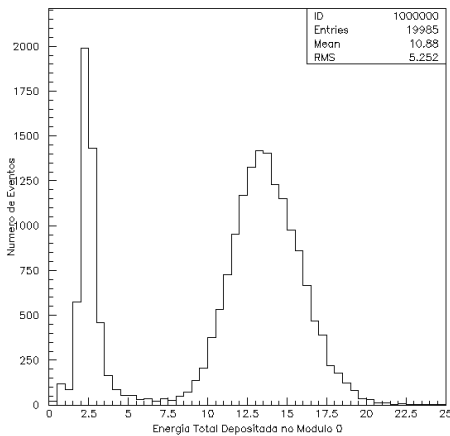


Figura 3: Deposição de energia no Módulo 0 causada por múons (valores próximos a zero) e píons (valores próximos a 20 GeV).

então a fase de aquisição de elétrons e, ao mesmo tempo (de novo *on-line*), o treinamento de uma segunda rede neural é feita com o feixe de elétrons contaminados por píons e múons, o conjunto de píons purificados na segunda fase e o conjunto de múons. Esta última rede possui três saídas, sendo cada uma associada a uma partícula, e deve fornecer um valor positivo (próximo de 1), caso esteja presente a partícula correspondente.

Para que se possa avaliar o desempenho desta metodologia e determinar um conjunto de parâmetros ótimo para o treinamento da rede, foi realizada uma simulação de todo o processo. Neste teste, a rede neural foi comparada com outras metodologias utilizadas classicamente na determinação desta separação. Estas metodologias se baseiam em cortes em certas variáveis, que dependem de fatores como energia ou posição do feixe. Vamos a uma rápida explicação desta metodologia clássica para que se possa compreender os testes realizados com a rede neural.

Como já referido anteriormente, o perfil de energia depositada no calorímetro revela a partícula incidente. Assim, sabe-se que os múons depositam pouquíssima energia no calorímetro, enquanto que píons e elétrons depositam uma quantidade de energia próxima ao valor nominal do feixe (por exemplo, 20 GeV), caso o protótipo de calorímetro em teste seja capaz de absorver completamente a energia destas partículas. Eventualmente, na prática, os módulos em teste são de dimensão reduzida, havendo alguma fuga de energia, por absoluta falta de material. A Figura 3 traz um histograma da deposição de energia num único módulo de calorímetro (46 células) para eventos de um feixe de píons. Entretanto, somente os eventos com energia depositada na faixa de 7 a 20 GeV poderiam representar píons na realidade. Os eventos de mais baixa energia que vemos na distribuição à esquerda, são os múons que contaminam este conjunto de píons. Pode-se então estabelecer um corte, por exemplo, em 7

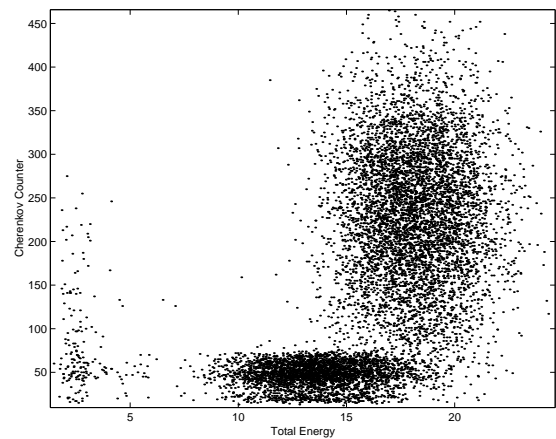


Figura 4: Energia total depositada (eixo das abscissas) contra o contador Cherenkov (eixo das ordenadas).

GeV para separar píons de múons.

Deve-se notar que há um certo vazamento de energia, pois o pico da distribuição de píons não se encontra ao redor de 20 GeV, o valor nominal de energia do feixe. Isto se dá por só termos utilizado, nos testes experimentais, metade de um dos 64 módulos que compõe o calorímetro final a ser instalado no ATLAS.

Para se separar elétrons de múons, o mesmo corte em energia pode ser realizado, entretanto, para se separar elétrons de píons a tarefa é um pouco mais complicada. Para se resolver esta dificuldade, um outro tipo de detector é colocado na linha do feixe, em períodos de teste. Este é o contador Cherenkov que mede a velocidade das partículas no interior de um meio formado por gases específicos [5]. Tipicamente, os elétrons passam bem mais rapidamente por este meio que os píons, fornecendo um outro possível corte. Na Figura 4 vemos eventos de um feixe de elétrons de 20 GeV, contaminado por píons e múons, a partir da leitura do contador Cherenkov. Os múons aparecem na região de baixa energia depositada no calorímetro. Na parte de mais alta energia (valores mais altos no eixo das abscissas), vemos que pode ser estabelecido um corte no eixo das ordenadas, relativo ao contador Cherenkov, separando os píons, parte inferior, dos elétrons (parte superior). Elétrons tendem a depositar plenamente a sua energia, pois desenvolvem cascatas mais estreitas e curtas. Desta maneira, o efeito de vazamento de energia é pouco significativo, ao contrário do caso dos píons.

De modo a avaliar a qualidade da discriminação das redes neurais, utilizamos a metodologia clássica, baseada nos cortes de energia depositada no calorímetro e no valor obtido do contador Cherenkov, como referência. Assim, como o método clássico é eficaz e se baseia em características conhecidas pelos físicos experimentais, ele irá fortalecer a confiança dos físicos no processamento neural mais sofisticado, caso haja um concordância significativa entre os métodos clássico e neural. Por outro lado, o método clássico é fortemente dependente da ener-

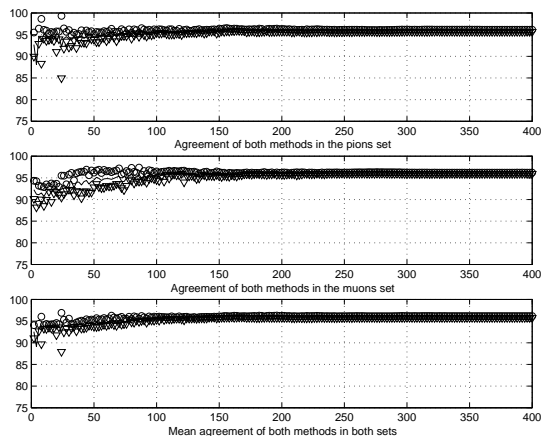


Figura 5: Concordância da rede neural com a metodologia clássica. Vê-se os valores máximo (círculos), mínimo (triângulos) e média (linha) obtidos por época de treinamento, a partir de 20 diferentes treinos.

gia, o que não acontece com o método neural [3]. Para o teste, com redes neurais, a mesma rede foi treinada com diferentes inicializações dos valores dos seus pesos e os eventos de treinamento também foram apresentados em ordens diferentes. Dessa forma, para um dado conjunto de parâmetros foi determinada a concordância obtida entre a rede neural e a metodologia clássica, após um certo número de passos, possibilitando-se que se tenha uma concordância mínima garantida. Desta forma, cada um dos possíveis parâmetros a serem estudados (razão de aprendizado, época, etc) foram sendo variados até que se determinasse um ponto ótimo, oferecendo um maior patamar de concordância. A partir daí outro parâmetro passou a ser escolhido e estudado.

Na Figura 5 temos os gráficos de concordância relativos a um total de 20 treinamentos da rede neural para a separação de píons e múons. A figura superior refere-se apenas aos eventos do conjunto de píons, a do meio aos eventos do conjunto de múons e a inferior diz respeito a ambos os conjuntos. No caso, a contaminação de múons no conjunto de píons estimada pelos métodos clássicos, era superior a 26%. Os símbolos marcados no gráfico são para cada época de treino: a maior concordância obtida (círculo), a menor (triângulo) e a média (contínua). Vemos que na fase final do treino a concordância é superior aos 95%, o que já representa um resultado positivo. O parâmetro que vínhamos estudando, neste caso, era a razão de decréscimo da razão de aprendizado [6] da rede. Assim, vários grupos de 20 treinamentos foram realizados, de modo que a única variação (além do início aleatório dos pesos e da seqüência de apresentação dos eventos) fosse a do parâmetro em questão.

A Figura 6 traz a evolução do mínimo valor de concordância obtido, para cada grupo de 20 treinos, após um certo número de passos de treinamento, em função da razão de decréscimo da razão de aprendizado. Vemos que a concordância no contaminado conjunto de píons (figura

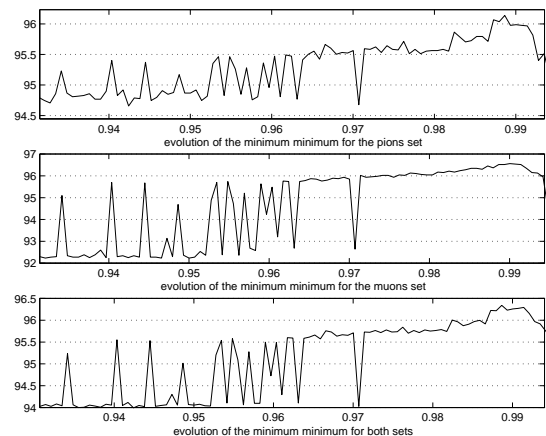


Figura 6: Evolução do nível de concordância em função da razão de decréscimo da razão de aprendizado. Vide texto.

superior) subiu de valores próximos a 95% para vir a superar a marca de 96% numa pequena faixa de variação de parâmetro. Na figura do meio, vemos que o nível de acerto para o conjunto de múons chega a se elevar mais de 4 pontos percentuais, atingindo também valores na faixa de 96%. Considerando o conjunto completo (figura inferior) conseguiu-se uma evolução de cerca de 2 pontos percentuais.

Concluídos os estudos dos parâmetros para esta fase do treinamento, a concordância total ficou acima de 96,4%. Utilizando a discriminação feita pela rede neural, fomos capazes de retirar a contaminação do conjunto de píons. A Figura 7 traz o resultado desta retirada. Vemos que a rede reconhece perfeitamente a contaminação, apesar da normalização utilizada sobre as entradas retirar a informação de valores totais de energia depositada. Portanto, a rede extrai as características fundamentais do perfil de deposição de energia das diferentes partículas.

Utilizando este conjunto de píons, identificados pela rede neural, passamos ao estudo da segunda rede, buscando os melhores parâmetros de treinamento para identificar a contaminação no conjunto de elétrons. Neste caso, a separação entre elétrons e píons, neste nível de energia relativamente baixo (de 20 GeV), é particularmente mais difícil [3].

A Figura 8 traz o estudo para a razão de decréscimo da razão de aprendizado, para o caso em que são incluídos os elétrons. Neste caso, foi incluído o contador Cherenkov para tornar eficaz a separação elétron-píon. Vemos que o nível de concordância mínima obtida chega a cerca de 92,5% no conjunto dos elétrons (para o qual havia mais de 37% de contaminação por píons e múons). No caso dos píons, a concordância fica também acima de 92% e, para os múons, acima de 97%. No total temos mais de 93,5% de concordância.

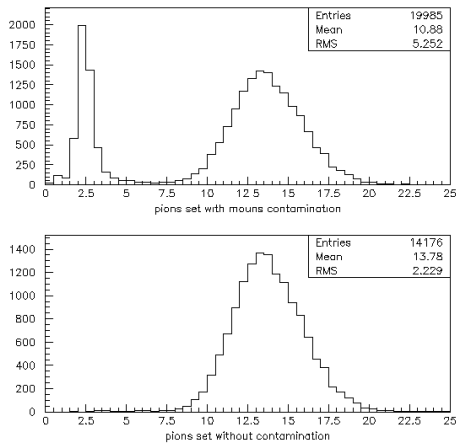


Figura 7: Retirada da contaminação do conjunto de píons pela rede neural. Conjunto original (superior), e os píons reconhecidos (inferior).

4. Simulação

A performance do discriminador neural nos encorajou a avançar na proposta de operação *on-line*, visando o desenvolvimento de uma rede neural que valide o que está trafegando no sistema de aquisição de dados, evitando que eventos indesejáveis sejam adquiridos. Como isso exige velocidade da aplicação, testes com uma implementação utilizando a tecnologia dos processadores digitais de sinais (DSPs) foram realizados.

O classificador neural foi implementado no ADSP-21062 (40MHz de clock) [7], sendo o DSP controlado por um PC (rodando Windows), que tem como função interfacear o DSP com o sistema de aquisição real, ou, no presente caso, simular esse sistema. A arquitetura interna deste DSP traz várias inovações que visam um aumento da velocidade de processamento. Entre os pontos importantes a se notar nesta arquitetura destacam-se :

- Uma memória SRAM que, conjuntamente com 4 barramentos de dados e instruções, permite múltiplos acessos, otimizando a velocidade de processamento.
- Um seqüenciador de programas completamente independente da Unidade Lógica e Aritmética, para não sobrecarregá-la com o cálculo de endereços de memória.
- Uma Unidade Lógica e Aritmética com 16 registradores, capaz de realizar diferentes tipos de operações em paralelo.
- Várias portas de acesso externo, que permitem a entrada e saída de dados de forma a prejudicar o mínimo possível o processamento.

A tarefa do sistema neural *on-line* é receber os eventos adquiridos pelo sistema de aquisição, treinar a rede neural discriminadora, atualizando o banco de dados acumulado de eventos físicos, e classificar os eventos recém

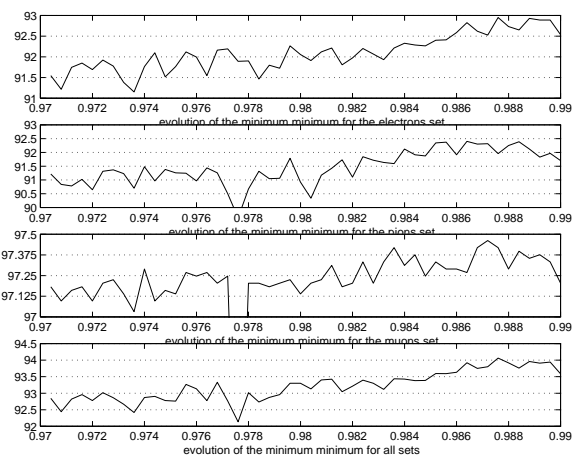


Figura 8: Evolução do nível concordância, em função da razão de decréscimo da razão de aprendizado, para o caso em que elétrons são incluídos. Vide texto.

adquiridos. Em testes experimentais com protótipos do calorímetro no CERN, o feixe experimental se apresenta em períodos de 14,4 segundos, estando ativo apenas durante cerca de 2 segundos e inativo no restante do tempo. Portanto a tarefa neural deve ser desempenhada em menos de 12 segundos.

Uma estimativa inicial da velocidade deste processamento, para uma codificação em linguagem C, indica um tempo de cerca de $250\mu s$ por operação de treino da rede neural, neste DSP. Os estudos simulados de operação indicam o treinamento da rede com uma época de 200 eventos. Por outro lado, apenas 100 passos de treinamento são suficientes para detectar-se corretamente a classe da partícula incidente. Desta maneira, seriam necessários cerca de 5 segundos para realizarmos o ciclo de treinamento, discriminando píons e múons, por exemplo, nesta tecnologia.

Além do treinamento, há necessidade de se converter a informação amostrada ao longo do tempo pelo conversor AD em valores de energia, para cada uma das 46 células que formam a informação de deposição de energia no calorímetro e calcular os correspondentes 23 valores médios geométricos. Considerando que a cada ciclo do feixe são produzidos cerca de 130 eventos, obtemos um tempo máximo de 1,3 segundos. O teste dos 130 eventos por ciclo ocorre em menos de 22 ms.

Considerando-se estas diferentes fases de processamento, a implementação do sistema em DSP é, portanto, viável. Além do mais, atualmente estamos avaliando o uso do ADSP-21160, da mesma família de processadores, que, graças a sua maior frequência de relógio (100MHz) e ao fato de possuir duas unidades lógica e aritmética, que podem trabalhar em paralelo, permitiria reduzir o tempo de processamento a um quinto do até aqui estimado, atingindo menos de 1,5 segundos.

Uma outra alternativa é se usar um PC ou um Power-PC. Resultados preliminares com um Pentium II

a 300MHz (em Linux) se mostraram bem promissores, sendo possível o processamento nos mesmos 1,5 segundos. A solução em Power-PC, numa máquina que faz parte do fluxo de dados do sistema de aquisição, ainda está sendo implementada.

5 Conclusões

Uma metodologia para identificação de partículas durante testes do calorímetro do ATLAS foi discutida. Dos resultados examinados se conclui que é possível se identificar corretamente tais partículas, mesmo quando temos contaminações de nível elevado (na faixa de até 40% nos casos estudados).

A metodologia de operação *on-line* se dividiu em três etapas: Na primeira é feita a aquisição de múons puros; na segunda temos a etapa de limpeza do conjunto de píons usando os múons puros já adquiridos; finalmente, na terceira etapa (utilizando o conjunto de píons purificado e o de múons) purifica-se o conjunto de elétrons.

Na primeira identificação de contaminação, em píons, foi obtido um nível mínimo de concordância de 96% entre a rede neural e a metodologia clássica, baseada nas características físicas das interação e na calorimetria. Na segunda descontaminação, no feixe de elétrons, este nível caiu para cerca de 93%. Esse valor mais baixo se deu bastante em função de ser mais complexo o processo de separação de partículas para níveis mais baixo de energia (20 GeV).

Uma solução baseada no DSP ADSP-21060 está sendo testada, codificando-se o processamento neural em linguagem C. Outras soluções, baseadas num PC (Linux) ou num Power-PC (Lynx-OS), estão sendo implementadas.

6 Agradecimentos

Agradecemos a colaboração ATLAS por disponibilizar os conjuntos de dados utilizados no presente trabalho e pelas importantes discussões, das quais resultaram muitas idéias empregadas nas soluções demonstradas.

Este trabalho recebeu o apoio das instituições CAPES e CNPq e também do CERN.

Referências

- [1] <http://www.cern.ch>.
- [2] D. O. Damazio and J. M. Seixas. Outsider identification in a neural particle discriminator based on calorimetry. In *Artificial Inteligency in High-Energy and Nuclear Physics*, 1999.
- [3] R. Wigmans. Advances in hadron calorimetry. *Rev. Nucl. Sci.*, 41:133–148, 1991.
- [4] S. Haykin. *Neural Networks: A Comprehensive Foundation, Second Edition*. Prentice-Hall, 1999.
- [5] ATLAS/The Tilecal Collaboration. *ATLAS - Tile Calorimeter Technical Design Report*. CERN/LHCC 96-42, 1996.
- [6] J. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1990.
- [7] *DSP/MSP Products Reference Manual*. Analog Devices Inc., 1995.