

**Proposta para abertura de concursos para contratação de docente em
“Informática de Materiais: modelagem e/ou experimental”**

Proponente: Grupo Teórico do Departamento de Física dos Materiais e Mecânica – IFUSP

- Antônio José Roque da Silva
- Caetano Rodrigues Miranda
- Gabriel Teixeira Landi
- Helena Maria Petrilli
- Lucy Vitória Credidio Assali
- Luís Gregório Dias da Silva
- Marília Junqueira Caldas

Área: Informática de Materiais

Temas: Cálculos de Primeiros Principios, Modelagem Molecular, Triagem de alto desempenho de materiais, Inteligência Artificial e Aprendizado de máquina aplicadas a Física da Matéria Condensada.

1. Justificar a adição de um docente na área;

O Instituto de Física da Universidade de São Paulo é um dos pioneiros no país na área de Física dos Materiais com reconhecimento internacional, em particular na frente de Cálculos de Estrutura Eletrônica e Modelagem Molecular para a caracterização e design de novos materiais. Na última década, considerável número de docentes (6) nesta área se aposentou, criando-se uma demanda para renovação do quadro em uma área ainda atual e pungente em Física e Ciência dos Materiais com forte apelo a aplicações no setor produtivo. Concomitantemente, a ciência dos materiais está sofrendo uma mudança de paradigma, devido a uma maior sinergia entre experimentos e simulações além da possibilidade de realizar-se uma grande quantidade cálculos computacionais. O status-quo anterior era se escolher alguns materiais de interesse e simulá-los a fim de entender-se melhor suas propriedades, o que produziu grandes avanços científicos. No entanto, a visão atual é escolher uma propriedade desejada, e varrer uma grande gama de materiais em torno das propriedades alvo. Este é um grande desafio, pois simular materiais requer uma compreensão profunda dos fundamentos e metodologia, sendo necessários estratégias específicas para cada tipo de material simulado. Contudo, o advento da área de informática de materiais (do inglês: materials informatics), pavimentou o caminho para tais simulações em massa, organizando as condições para simulações serem realizadas de uma maneira consistente e reprodutível. Assim, interpretar essa grande quantidade de dados obtidos é o desafio atual da ciência de materiais, uma vez que tendências podem ser estabelecidas, mas podem ser altamente complexas devido a vasta variabilidade dos sistemas estudados. Para tanto, espera-se que além dos conhecimentos fundamentais Física e Química, o docente tenha um perfil híbrido orientado a Ciência de Dados, em particular, Métodos de Inteligência Artificial e Aprendizado de Máquina.

2. Impacto da contratação no âmbito do Instituto de Física;

O Brasil não tem iniciativas organizadas na linha de Materials Informatics, além de não existirem pesquisadores especializados neste tipo de pesquisa no Instituto de Física da USP (IF-USP). Portanto, é importante que o IF-USP tenha docentes desta área, para estar na vanguarda da ciência de materiais, formando recursos humanos e se estabelecendo como líder no Brasil e com o objetivo de organizar uma grande iniciativa nesta área a nível nacional, levando a uma projeção internacional do IF-USP na área. Também importante realçar a mudança de paradigma na graduação e pós-graduação do IFUSP onde espera-se fortalecer conhecimentos em Física Computacional e modelagem numérica de sistemas físicos, havendo uma necessidade latente quanto a docentes com conhecimentos em Ciência de Dados, em particular Aprendizado de Máquina.

A contratação poderá beneficiar diretamente os grupos teóricos e experimentais na área de Materiais e Nanociências que permeiam diversos departamentos do IFUSP.

3. Relevância atual da área (nacional e internacionalmente);

A área de informática dos materiais é recente englobando a ciência dos materiais, físico-química, física aplicada, nanociências, física da matéria condensada, entre outras. Este nome foi recentemente consolidado a partir de uma série de iniciativas de organizar e ampliar o conhecimento sobre materiais. A primeira grande iniciativa nesta direção foi a “*Materials Genome Initiative*” (MGI), anunciada em 2011 por Barack Obama, então presidente dos Estados Unidos da América. Seu objetivo é acelerar a descoberta, design e aplicação de novos materiais com uma combinando sinergicamente experimento, teoria e computação de maneira integrada e com alto rendimento (alto rendimento vem de *high-throughput*, que consiste na análise sistemática de uma grande quantidade de materiais).

A MGI se desdobrou em uma série de projetos: “*Designing Materials to Revolutionize and Engineer our Future*”, “*Lightweight Innovations for Tomorrow*”, “*Energy Materials Network*”, “*virtual, high-throughput experimentation facility*” e também o “*Materials Project*”. Alguns projetos tem viés experimental, nos quais experimentos são realizados em massa. Outros tem como foco simulações computacionais também realizadas em massa. A proposta envolve utilizar os resultados da grande quantidade de simulações realizadas, para uma triagem prévia de materiais para a aplicação desejada. Então, experimentos podem ser realizados para testar a viabilidade dos materiais propostos, levando então a possíveis candidatos de aplicações industriais. Além disso, outros países também tem projetos a nível nacional na mesma linha, entre os principais estão o programa Suíço NCCR MARVEL (*Materials’ Revolution: Computational Design and Discovery of Novel Materials*), outro programa americano o AFLOW (*Automatic FLOW for materials discovery*) e o programa alemão NOMAD Laboratory (*The Novel Materials Discovery Laboratory*).

De todas as iniciativas de simulações computacionais em massa, é necessário que os seus resultados sejam organizados de maneira sistemática e tornem fácil a reprodutibilidade. Isso já é uma realidade para resultados experimentais, com exemplo das bases de dados cristalográficas, que são as mais variadas, como a The Inorganic Crystal Structure Database (ICSD), Crystallography Open Database (COD), Cambridge Structural Database (CSD), the Protein Data Bank (PDB), entre outras. Inspirados nessas iniciativas, aliada a necessidade de sistematizar os resultados, foram criadas bases de dados para cálculos teóricos. Tais bases de

dados advém também das iniciativas já citadas, como a *Materials Cloud* do NCCR MARVEL, já com mais de 7 milhões de estruturas com cálculos de teoria do funcional da densidade (DFT do inglês *Density Functional Theory*) associados; assim como a AFLOW com mais de 3 milhões de estruturas calculadas; também a Materials Project, do projeto homônimo, com perto de um milhão de cálculos DFT associados; o projeto NOMAD também possui um repositório de dados, com 5 milhões de entradas. Além das bases de dados associadas a diferentes projetos, há uma iniciativa de um banco de dados aberto, o *Open Quantum Materials Database*, que conta com mais de 600 mil materiais. Nessas bases de dados os pesquisadores submetem os mais variados resultados, como parâmetros de rede e coordenadas atômicas relaxadas, módulos de elasticidade, estrutura de bandas, espectro vibracional, entre outras. Além das propriedades, as entradas e saídas dos cálculos são armazenadas, tornando extremamente simples o processo de reprodutibilidade.

Como a informática de materiais é uma área de fronteira, as publicações produzidas nesta área são de alto impacto, como exemplo, apenas o *Materials Project* e o *NCCR Marvel*, já publicaram 200 artigos nas seguintes revistas: Nature, Nature Chemistry, Nature Communications, Nature Materials, Nature Nanotechnology, Nature Physics, Nature Reviews Materials, Physical Review Letters, Proceedings of the National Academy of Sciences of the United States of America, Science. Na Figura 1, são mostradas as publicações com maior impacto dos dois projetos, entre as 208 publicações, 36 são no Physical Review Letters, uma das mais importantes revistas da área de física. Também ao observarmos a Figura 2, onde apenas as revistas com o maior número de publicações de ambos projetos é mostrada, nota-se que o carro chefe das publicações é na Physical Review B, uma das mais importantes na área de física da matéria condensada. Desta maneira, notamos que a física da matéria condensada é o principal instrumento da *Materials Informatics*, onde a mudança de paradigma das simulações computacionais e também experimentos está ocorrendo rapidamente.

4. Prognóstico de potenciais candidatos;

Jovens lideranças vem se formando na área já com um portfólio de publicações expressivo em grupos renomados dentro da própria USP, Unicamp, UFABC, UFMG e UFRJ, que usualmente estão migrando para o exterior por falta de opção nacional. Este perfil também tem o potencial de atrair candidatos internacionais dentro das várias frentes do Materials Project, NCCR Marvel, NOMAD, AFLOW, entre outros. O grupo de docentes do IFUSP mantém colaborações de longa data com os pioneiros dessa área, tais como Gerbrand Ceder e Kristin Persson (Materials Project), Nicola Marzari (NCCR Marvel), Mathias Scheffler e Claudia Draxl (NOMAD).

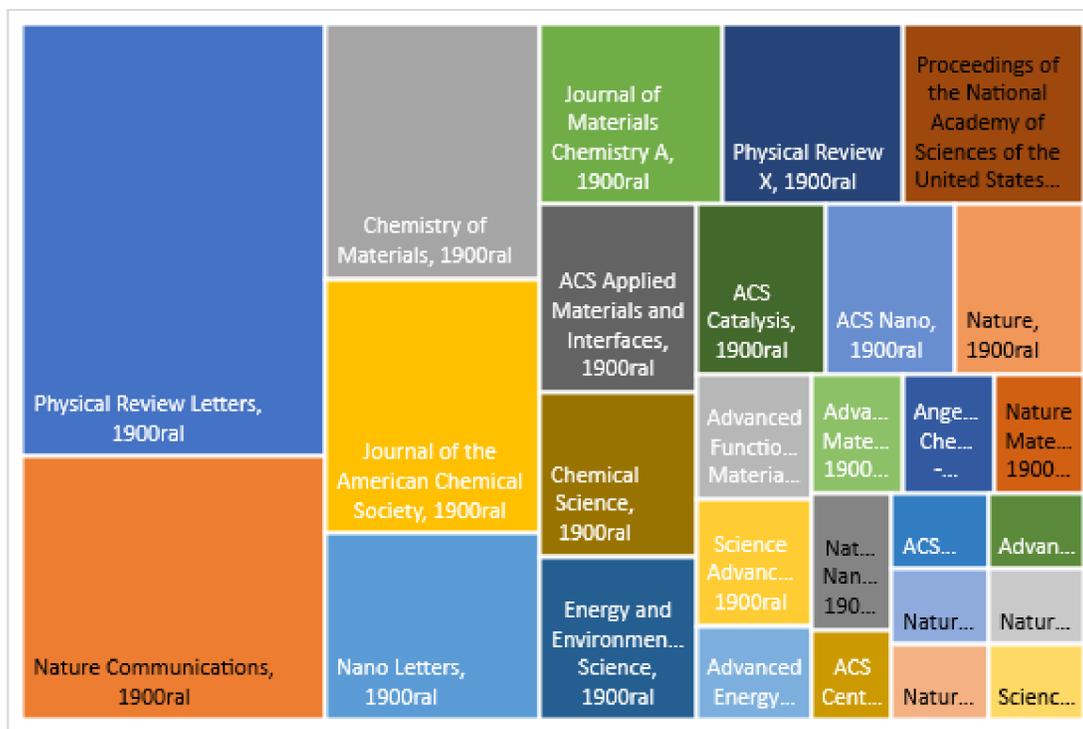


Figura 1: Publicações combinadas do Materials Project e NCCR Marvel apenas em revistas com parâmetro de impacto maior que 8.

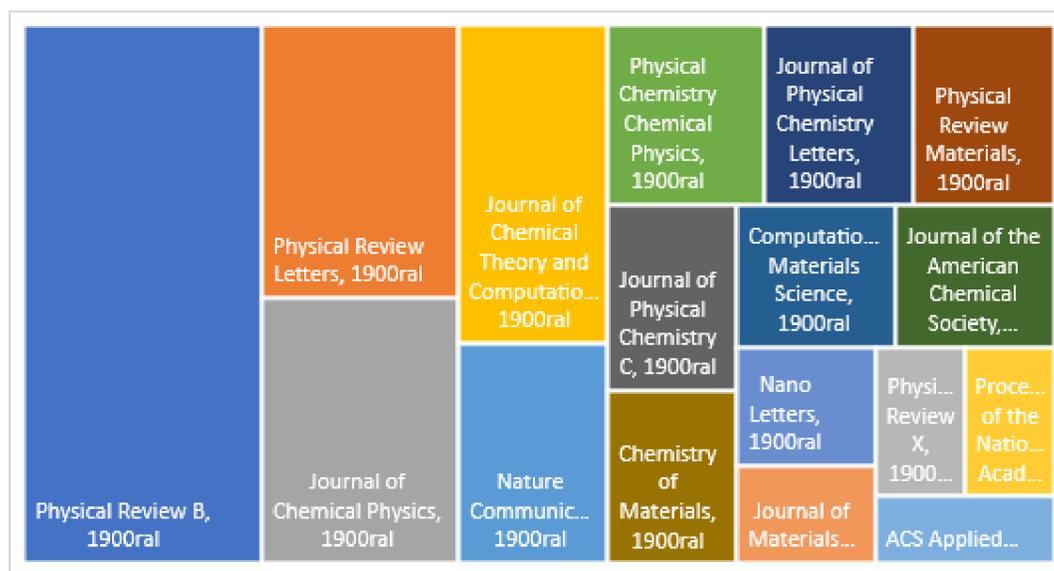


Figura 2: Revistas com maior número de publicações do Materials Project e NCCR Marvel.

5. Viabilidade da execução de projetos na área.

Esta área demanda uma grande quantidade de recursos computacionais, mas, tem a vantagem de poder usar recursos espalhados. Isso porque, a ideia é enviar os dados para simulações para as diferentes facilidades de computação de alta performance, executar os cálculos nos sistemas de filas locais, então coletar os resultados e organizar em uma base de dados local do grupo de pesquisa. Assim, facilidades públicas nacionais podem ser utilizadas, como o SINAPAD, que tem vários centros computacionais espalhados pelo país, incluso o Santos Dumont do LNCC, que teve recente upgrade, contando agora com uma capacidade computacional de 5.1 Petaflops/s. Em particular, a USP mantém um setor específico para computação de alto desempenho, disponibilizando recursos computacionais aos docentes através da InterNuvem USP. Ainda, existe também o SENAI-SIMATEC que dispõe de uma grande capacidade computacional, atrelado a empresas do setor produtivo.

Utilizar-se esta vasta gama de recursos computacionais espalhados pelo país é algo difícil de ser realizado manualmente. No entanto, com as ferramentas de *Materials Informatics* isso torna-se viável, diminuindo o investimento necessário para começar o trabalho na área.

Também é importante ressaltar que o IFUSP também já conta com uma ótima infraestrutura disponível no CCIFUSP para receber sistemas de computação de alta performance, desta maneira, torna fácil a instalação de novas máquinas com recursos obtidos pelo pesquisador via agências de fomento ou convênios com a indústria.