Accelerating LQCD Calculations Using the Tiramisu Compiler

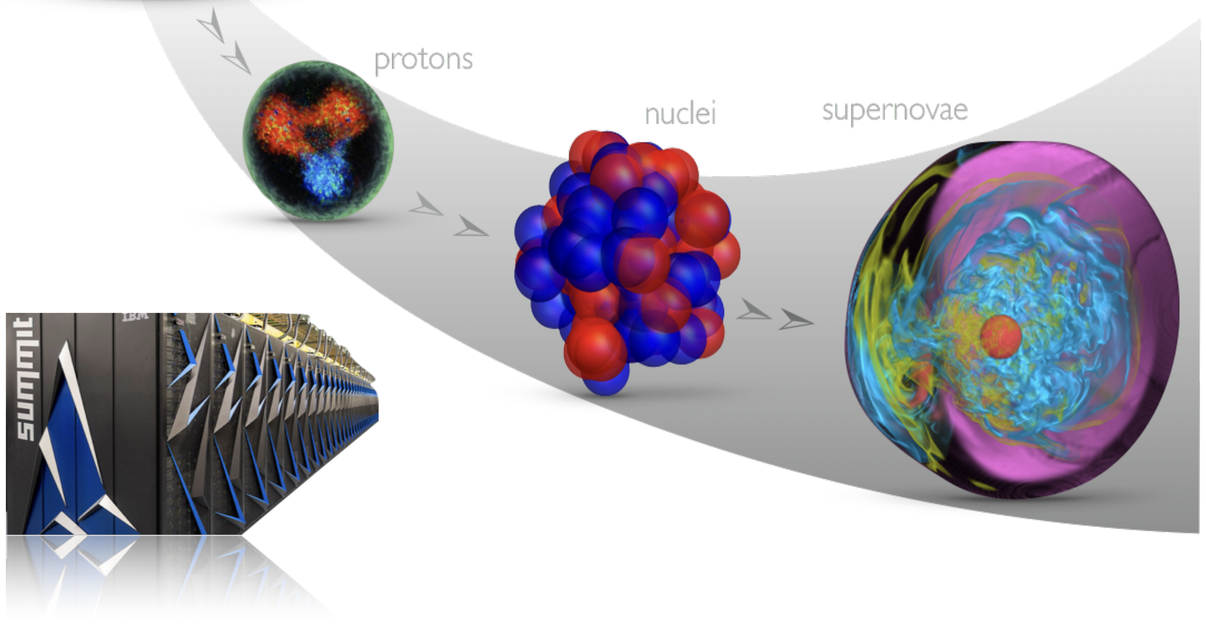
**Riyadh Baghdadi, Michael Wagman,
Andrew Pochinsky, Saman Amarasinghe, William Detmold**

Massachusetts Institute of Technology

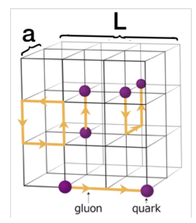
- The Standard Model of particle physics combines the strong, weak and electromagnetic forces

- The strong force, mathematically described by Quantum Chromodynamics (QCD), binds quarks and gluons together into protons and nuclei

- Strong interactions are critical to understanding the structure of the proton and nuclei, the expansion of the universe, searches for dark matter and physics beyond the Standard Model

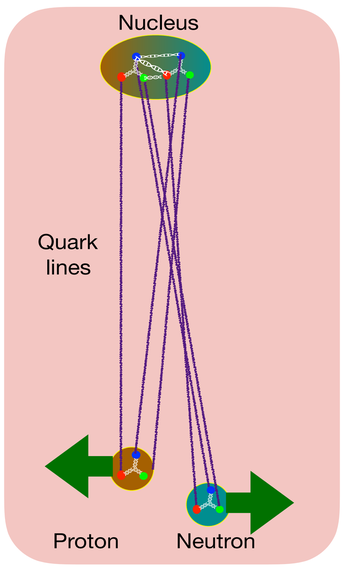


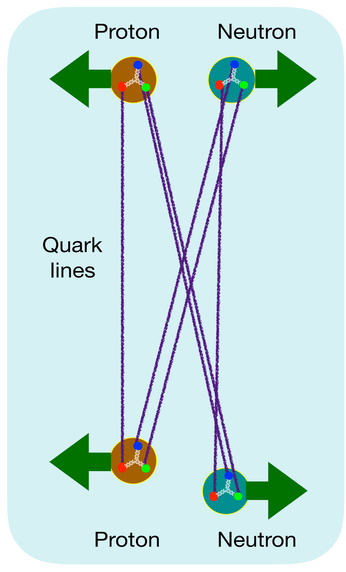
- To solve QCD in a controlled manner requires the numerical method of **lattice QCD**



- Discretise and compactify spacetime, solve field equations on 4D spacetime lattice using Metropolis style algorithms

- Simple aspects of proton - neutron scattering have been studied in lattice QCD using simulations like the red cartoon





- Simulations like the blue cartoon provide greater statistical precision and greater control over the proton-neutron scattering state, but are more computationally demanding

The Tiramisu Compiler

Generate fast code

Target different architectures

Target Architectures

Distribution/Communication (MPI)

X86 CPUs

GPUs (CUDA)

Tiramisu Highlights

- A simple C++ API to express tensor operations

- Large set of optimizations: multicore parallelization, vectorization, blocking, operation fusion, data transformations, loop reordering, ...

- Manual code optimization

- Soon: automatic code optimization, integration in NumPy

Pseudocode

```
for (i=0; i<100; i++)
   for (j=0; j<100; j++)
      C(i,j) = 0;
```

Tiramisu Code

```
// Tiramisu algorithm
var i("i", 0, 100), j("j", 0, 100);
computation C({i,j}, 0);

// Optimizations
C.parallelize(i);
C.vectorize(j, 4);
```

Evaluation on Multicore CPU

Tiramisu VS Reference C Implementation (Auto-vectorized)

130x speedup

Optimizations:
- Vectorized (AVX2)
- Parallelized (24-cores, single-node)
- Full fusion
- Loop reordering
- Data layout optimization

Convolution - Tiramisu VS Intel MKL (CPU)