

# Semistochastic importance sampling of second-quantized operators in determinant space: application to multireference perturbation theory

Adam A Holmes

June 1, 2021

## Abstract

We present an algorithm for semistochastically applying a second-quantized operator to a Slater determinant using importance sampling. Our method efficiently applies the largest-magnitude terms deterministically, and enables sampling of the remaining, smaller terms, with probability proportional to a function of the terms' magnitudes. The time complexity of the deterministic component scales as only the number of large-magnitude terms (rather than all terms), as in the heat-bath configuration interaction (HCI) paper, and the time complexity of each sample in the stochastic component scales only as the logarithm of the number of orbitals. We then use our algorithm to perform efficient, semistochastic multireference perturbation theory in semistochastic heat-bath configuration interaction (SHCI).

## 1 Introduction

Many quantum chemistry algorithms operate in the space of Slater determinants. These methods include deterministic algorithms such as selected configuration interaction, stochastic/semistochastic algorithms using classical statistics such as semistochastic multireference perturbation theory, and projector Monte Carlo algorithms such as full configuration interaction quantum Monte Carlo (FCIQMC).

In all of these determinant-space algorithms, one of the key steps is applying a second-quantized operator (usually the Hamiltonian) to a Slater determinant (or linear combination of them):

- In selected CI, a variational wavefunction is obtained iteratively by ‘selecting’ new determinants each iteration using a criterion that is a function of the Hamiltonian times the previous iteration’s wavefunction. Once

the new determinants are selected, their coefficients are variationally optimized using an algorithm such as Davidson, in which the Hamiltonian is diagonalized in a space of Krylov vectors, each of which is a function of the Hamiltonian times other Krylov vectors.

- In semistochastic multireference perturbation theory as used in SHCI, contributions to the perturbation theory expression are computed by applying the Hamiltonian to determinants sampled from the variational Hamiltonian.
- In FCIQMC, the power method is simulated in the full set of Slater determinants by repeatedly stochastically (or semistochastically) applying a projector operator – a function of the Hamiltonian – to a stochastic representation of the wavefunction. The energy is estimated using a mixed energy estimator, which makes use of the Hamiltonian times a trial wavefunction in determinant space.

Since the application of the Hamiltonian to a Slater determinant is such a key step in all of these algorithms, we explore methods to evaluate it. Broadly, there are three main approaches:

1. **Deterministic:** Simply evaluating all single and double excitations from the initial Slater determinant. Time complexity:  $\mathcal{O}(N^2M^2)$ , space complexity:  $\mathcal{O}(N^2M^2)$ .
2. **Stochastic:** Sampling excitations according to some distribution. Reduces the time and space complexity, but introduces a stochastic uncertainty (and in projector methods, the fermion sign problem!).
3. **Semistochastic:** Evaluating the largest-magnitude components, and sampling the remaining, smaller components. Has a much reduced time and space complexity relative to the deterministic approach, but with a greatly reduced stochastic uncertainty relative to fully stochastic methods. In projector methods, it also mitigates the bias due to the fermion sign problem.

It should be noted that a simple version of semistochastic importance sampling has been used before in the context of FCIQMC. However, in that case, the deterministic component was pre-computed, and the stochastic component was not disjoint from it: there was some probability of sampling an excitation that already existed in the deterministic component and therefore had to be discarded.

In this paper, we describe a unified approach to semistochastic importance sampling, in which the deterministic component is computed efficiently on the fly, and the stochastic component efficiently samples only the remaining terms left out of the deterministic component.

