Quantum Simulation of Asphalt Binder Aging – Phase 2

Team Name: Hydrodynamic **Email:** venkatn93@gmail.com

Team Member: Niraj Venkat **Job Title & Organization:** Teaching Assistant at MITx

1. Molecular System Selection & Observable Definition:

For our representative molecule involved in asphalt aging we have selected C_8H_6S, named **benzothiophene (BT)**. It belongs to the **aromatics** class within the **maltenes** fraction of asphalt. This classification is critical because the aromatic maltenes are mobile and more reactive, especially under oxidative conditions. We choose BT because it balances chemical relevance (sulfur-containing PAH) with computational feasibility for developing and showcasing the quantum simulation pipeline.

To get comfortable with our quantum-chemistry toolkit, we began by computing BT's ground-state energy. We now aim to probe the key aging pathways—sulfoxide and carbonyl formation—and our goal is to deliver the corresponding oxidation energies for each.

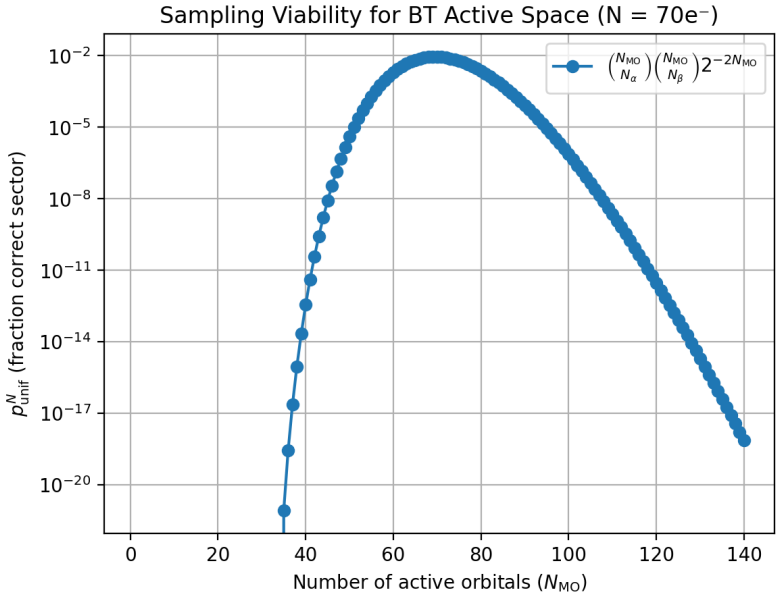

Figure 1 - Optimizing BT molecular orbital sampling

BT—with restricted Hartree-Fock energy of -703.79 Ha—consists of 70 electrons (35 α/35 β) in 97 molecular orbitals; the HOMO is 34 and the LUMO 35. We created settings specifically optimized for our QPU-CPU pipeline, as well as future hardware. `OrbitalSetting.MEDIUM`—which is currently used for ground-state energy calculation—freezes core orbitals 0–25 defining an active space of 16 electrons in 16 orbitals (26–42).

As a first pass, we look at the uniform-sampling probability over electronic sectors, which is shown in Figure 1. We plot the mean probability that a bitstring sampled from a uniform distribution would land in our electronic subspace. We then perform a slope analysis of the curve based on our desired level of accuracy. We estimate that for BT, the natural number of molecular orbitals to simulate is 42, requiring a qubit count of 95 (84 spin-orbitals + 11 ancillas), which is still prohibitively expensive. In comparison, the largest quantum chemistry experiment conducted by IBM was (54e, 36o) for Fe_4S_4 which required 77 qubits total including ancillas, and 6400 nodes of the Fugaku supercomputer running in parallel. We think that a near-term approach for **simulating BT is challenging, but feasible**.

2. Initial Hamiltonian Construction & Encoding Strategy:

We import the molecular description of BT from PubChem and choose a basis set. We assume that BT has no symmetries.

For the task of **probing oxidation energies**, we plan to consider the relevant charge and spin as well. Our current approach involves positioning O near a specific site of BT (either S or C) and computing the energy of the resulting complex. We plot the energies for various distances of approach and because each geometry requires a fresh energy evaluation, all data is generated with classical `PySCF` methods. We have performed these scans using both molecular O_2 and the O• radical to capture different oxidation pathways. Our results do not show a conclusive signature of the oxidation energy, and instead we showcase our ground-state energy analysis of BT alone that leverages quantum computing.

Once the HF state is prepared on the quantum computer, we employ a truncated version of the local unitary cluster Jastrow (LUCJ) ansatz. The HF state and LUCJ operator are mapped onto qubits using the Jordan-Wigner (JW) transform. SQD post-processing requires that we sample bitstrings in the occupation number (Fock) basis. **JW has its advantages**—occupation numbers are local which simplifies orbital rotations (used in LUCJ) and Fock basis sampling has no overhead.

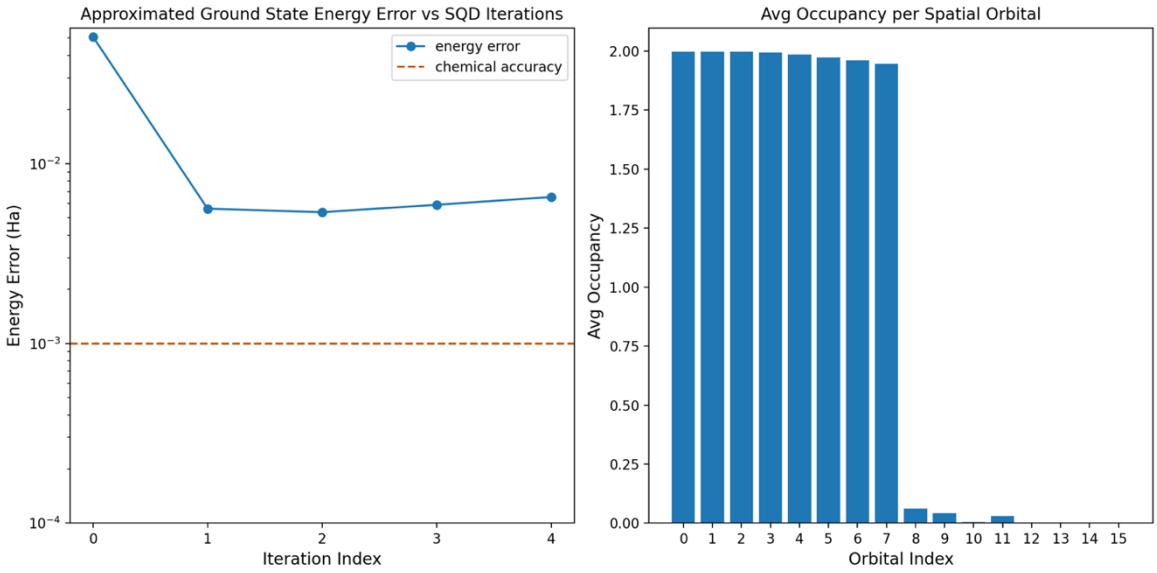

Figure 2 – Results of the SQD workflow for BT ground-state estimation

By performing various optimizations (e.g. the eigensolver used in self-consistent configuration recovery is now being run in multithreaded dispatch), we were able to reach within an order of magnitude of chemical accuracy. Given that **we currently use the minimal STO-3G basis**, we can further improve accuracy by switching to the systematically convergent cc-PVnZ basis sets. They use additional unoccupied ('virtual') orbitals which are essential for capturing the missing correlation energy. cc-PVnZ is therefore critical for reliably modeling BT oxidation pathways (and, more broadly, the chemistry of larger asphalt fragments).

3. Quantum Algorithm Development Plan:

We are running the sample-based quantum diagonalization (SQD) workflow (see Figure 2), where the quantum sampling is done on IBM Quantum (Eagle r3), and classical post-processing on a 12 core Macbook. More expressive basis sets that demand larger active spaces increase the qubit count linearly, while the classical space-time cost increases exponentially. Quantum error correction (QEC) will lower the number of samples needed to obtain good bitstrings. Azure Quantum Resource Estimator suggests that QEC, if deployed, will significantly increase the number of physical qubits needed.

The `PySCF` library is used throughout our project, to construct the molecular Hamiltonian as well as perform classical quantum chemistry simulations. We compute approximations to the ground-state energy with `PySCF` and have implemented the following settings which will change the code path to use a combination of the RHF, MP2, DMRG, CASCI, and CCSD calculations:

```
class OrbitalSetting(Enum): # This will impact CPU and QPU resources needed
    LOW = 12
    MEDIUM = 16
    HIGH = 23 # uses 52 qubits on Eagle, see doi.org/10.1021/acs.jctc.5c00075
    ULTRA = 26
class ActiveSpaceSetting(Enum): # This will impact convergence to chemical accuracy
    NATURAL_MO = auto()
    HOMO_LUMO = auto()
class PostHFSetting(Enum): # This will select the post-HF method used
    CASCI = auto()
    DMRG = auto()
```

RHF is run in every code path for the initial mean-field approximation to the ground-state energy, which will be improved upon in later steps. `ActiveSpaceSetting.HOMO_LUMO` gives us an active-space selection that depends on `OrbitalSetting`. We then either run CASCI or DMRG to calculate the reference energy labelled as chemical accuracy. CCSD is used separately from other methods only to calculate t_1 and t_2 amplitudes for the LUCJ ansatz. While accurate

at equilibrium, CCSD starts to overestimate and then overcorrelates especially in the challenging multi-reference regime of oxidation transition states.

We are **running the eigensolver multithreaded** over 12 CPU cores, which significantly boosts performance over the single-threaded loop. This allows us to scale to 24 batches which is putting us within reach of chemical accuracy. Another key parameter for each batch (indexed by k) is the number of samples per batch, which provides an upper bound on the dimension d of the subspace $\mathcal{S}^{(k)}$ into which the many-body Hamiltonian is projected. **Increasing samples per batch** gives us a high chance of getting the ground state in our subspace that we classically diagonalize. If the sampled bitstrings have the incorrect particle number/Hamming weight, this is where configuration recovery will attempt to fix things.

Drawbacks of other approaches to quantum chemistry:

- Quantum Phase Estimation (QPE): Requires deep quantum circuits with controlled unitaries which are expensive on NISQ devices without all-to-all connectivity. In the fault-tolerant regime, QPE is guaranteed to converge if we have prepared a state having polynomial overlap with the ground state. In the NISQ regime we see limitations with certain Hamiltonians, for example, ones with vanishingly small energy gap where one needs extremely fine phase resolution.
- Variational Quantum Eigensolver (VQE): No convergence guarantees, incurs a prohibitive number of measurements. VQE suffers from barren plateaus, which occur even in noiseless simulations due to fundamental theoretical reasons.
- Sample-based Krylov Quantum Diagonalization (SKQD): SQD on samples from Krylov states $|\psi_k\rangle := e^{-ikH\Delta t}|\psi_0\rangle$. To keep low circuit depths, one must address lattice problems which are amenable to being mapped easily using the layout of the quantum processor. As a consequence, chemistry use cases are out of reach for SKQD in the near-term.

4. Execution Platform, Feasibility and Roadmap:

Classical Reqs.: Linux 4× x86 CPU/v100 GPU—leverage parallelism for reference energy calc. & SQD post-processing

Quantum Reqs.: IBM Heron r1/r2 QPU—tunable couplers & better signals to capture asphalt binder oxidation chemistry

Justification for Hardware Access and Future Work:

- To study oxidation dynamics, we have explored various spin-unrestricted SCF, DFT, and CI methods available in `PySCF`. Currently, we face numerical stability and convergence issues. Our focus is on generating a potential energy curve with a clean elbow—a hallmark of oxidation energy. Using `GPU4PySCF` acceleration is likely to improve performance, enabling faster energy calculation along the reaction coordinates.
- LUCJ ansatz aligns the α/β spin orbitals along the main diagonal of the device. As we scale LUCJ + SQD, the QPU-CPU pipeline faces greater resource demands, see Figure 3 and Figure 4. We mainly require Heron for lower 2Q error rates and 200K+ CLOPS.
- We may consider some other fermion-to-qubit mappings like parity encoding and Bravyi-Kitaev transform, which could potentially improve circuit depth when encoding multi-qubit gates like those in the LUCJ operator.
- Under investigation is the usage of `qiskit.quantum_info.Z2Symmetries` for qubit tapering by taking advantage of the particle number conservation and spin $\mathbb{Z}_2$ symmetries of the Hamiltonian. Experiments show a two-qubit reduction is possible.
- An alternative active-space selection is being considered which we call `ActiveSpaceSetting.NATURAL_MO`. It ranks orbitals according to their natural occupation numbers and incorporates those with the greatest deviation from full occupation—which should capture key correlation effects in BT's ground state and along its sulfoxide/carbonyl oxidation pathways.
- We require access to a Linux system to run HCI calculations on `DICE`, which is the reference classical method used by IBM. As indicated by the SQD experiments on N_2 triple bond dissociation, HCI improves chemical accuracy especially when switching to the cc-pVDZ basis set. We can try `qiskit_addon_dice_solver` which calls the internal eigensolver used in `DICE` instead of the iterative Davidson method in `PySCF`.

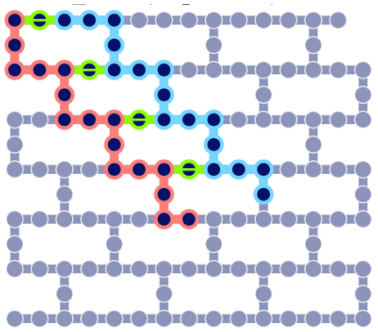

Figure 3 - Current mapping of BT:
Active space: 16 orbitals
Qubit count: 32 + 4 = 36

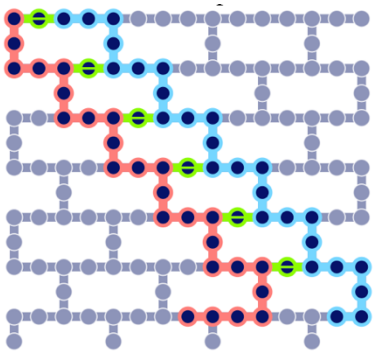

Figure 4 - Mapping BT on Heron r1:
Active space: 26 orbitals
Qubit count: 52 + 6 = 58