

ARTICLE

HERO (Hessian-Engineered Relaxation Optimizer): Suppressing “Hessian Pollution” for Accelerated First-Principles Structural Relaxation

Mingzhe Li^{1,2,3,4}, Piao Ma^{2,4}, Limin Li⁴, Weijie Yang^{1,*} and Hao Li^{3,*}

¹Department of Power Engineering, North China Electric Power University, Baoding, China

²Gusu Laboratory of Materials, Suzhou, China

³Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai, Miyagi, Japan

⁴Suzhou MatSource Technology Co., Ltd., Suzhou, China

*Corresponding Authors: Weijie Yang, Email: yangwj@ncepu.edu.cn; Hao Li, Email: li.hao.b8@tohoku.ac.jp

Received: 15 January 2026; Accepted: 02 April 2026; Published: 08 May 2026

ABSTRACT: Structural optimization is a fundamental step in density functional theory (DFT) calculations, typically driven by the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimizer. However, the standard BFGS algorithm relies on a local quadratic approximation of the potential energy surface (PES), which frequently breaks down in highly non-quadratic regimes typical of complex surface adsorption systems and defective bulk materials. This breakdown leads to “Hessian pollution”, a phenomenon where higher-order anharmonicities introduce spurious off-diagonal inter-atomic couplings that distort curvature estimates and significantly stall convergence. Herein, we propose a physics-inspired algorithmic intervention to the BFGS method that systematically suppresses this pollution. Once the maximum residual force drops below a specific activation threshold (e.g., 0.5 or 0.1 eV/Å), our approach conditionally resets all off-diagonal Hessian blocks, and introduces an isotropic background stiffness strategy where these blocks can be repopulated with a small positive constant rather than zeroed completely. This balances the robust stability of diagonal dominance with accelerated convergence speed. Implemented as an add-on to the Atomic Simulation Environment (ASE) Library, the method is lightweight, transferable, and compatible with standard DFT codes. Tests across diverse chemical systems, including atomic and molecular adsorbates (O*, H*, CO*) on Pt(111) surfaces and defective bulk oxides (WO_{3-x}), demonstrate substantial reductions in the number of required force calls without biasing the final optimized geometry. It offers a practical tool for high-throughput DFT workflows that eliminates the need for domain-specific training. This method is available via our open-source package, Hessian-Engineered Relaxation Optimizer (HERO).

KEYWORDS: Hessian-engineered relaxation optimizer; density functional theory (DFT); structural optimization; BFGS algorithm; hessian pollution; structural relaxation

1 Introduction

Structural optimization is a fundamental and essential step in density functional theory (DFT) calculations [1,2], providing the equilibrium geometries required for accurate computations of energetics, electronic structures, and physical and chemical properties [3]. Among the optimization algorithms, quasi-Newton methods, particularly the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [4] have become the standard choice due to their favorable balance between efficiency and robustness. By iteratively updating an approximate Hessian or its inverse, BFGS circumvents the need for costly explicit second-derivative calculations while providing effective curvature information for guiding ionic relaxations [5].

Despite its popularity, the BFGS method is not immune to numerical instabilities. Improper definition of the search direction due to an inappropriate initial Hessian approximation or an ill-conditioned Hessian can hinder convergence [6]. Its formulation relies on an implicit assumption: the local potential energy surface (PES) [7] can be approximated by a quadratic function. In this idealized case, the gradient difference is strictly linear, meaning the secant condition perfectly captures the constant local curvature, and the update progressively converges to the true Hessian. However, real PESs encountered in DFT calculations are intrinsically highly non-quadratic, especially in surface and adsorption systems where strong anharmonicity and many-body couplings are present. As a result, the gradient difference vector γ_k inevitably incorporates higher-order contributions, which contaminate the Hessian approximation [8]. This “Hessian pollution” distorts the curvature estimate, producing ill-conditioned matrices and spurious couplings that slow down convergence.

Several approaches have been developed to mitigate stability issues in structural relaxation. Strategies such as preconditioning schemes [9], trust-region methods, and machine-learned surrogate models [10,11] have been widely employed to improve convergence and stability. Recently, the field has seen rapid advancements [12] with the introduction of universal machine learning interatomic potentials (MLIPs), such as MACE [13] and CHGNet [11], which offer near-quantum accuracy for geometry optimization at a fraction of the computational cost [14]. Furthermore, novel batched optimization frameworks and comprehensive benchmarking studies have been proposed to evaluate and accelerate MLIP-driven structural relaxation across diverse energy landscapes [15,16]. Nevertheless, these methods often impose constraints: surrogate models typically require extensive domain-specific training datasets or incur significant on-the-fly computational costs [10,11], while advanced optimization schemes may necessitate intricate parameter tuning to match specific system characteristics. In high-throughput computational workflows [17], where thousands of structural relaxations must be performed efficiently, a lightweight and transferable solution is particularly desirable.

Modern advancements that address these limitations include the FIRE algorithm [18] for robust dynamics-based minimization, the Universal Exponential Preconditioner [9] for low-overhead geometric scaling, the TRIC system [19] for parameter-free coordinate transformations, the FLARE [10] active learning framework for on-the-fly surrogate training, and CHGNet [11] as a pretrained universal potential for rapid pre-relaxation. However, these methods have specific trade-offs: the FIRE algorithm may require more iterations than quasi-Newton methods in harmonic regions [18], the exponential preconditioner is less effective for complex molecular systems without hybridization [9], and machine learning approaches like CHGNet or FLARE often still require final *ab initio* refinement to ensure high-precision accuracy in novel chemical spaces [10,11,20].

Herein, we propose a physics-inspired modification to the BFGS algorithm that systematically suppresses Hessian pollution in late optimization stages. The method exploits the block structure of the Hessian in atomistic systems: once the residual forces fall below a given threshold, all off-diagonal blocks are reset to either zero or a small positive constant. This block-diagonalization eliminates spurious interatomic couplings introduced by higher-order nonlinearities, restoring stability by physically decoupling the coordinates and purging the accumulated Hessian pollution. Furthermore, we introduce a dynamic structured Hessian preconditioning strategy by injecting an isotropic background stiffness (an empirical positive coupling constant) into the off-diagonal blocks to a small positive constant. This mitigates the conservative nature of a purely diagonal Hessian, balancing robustness with convergence speed. The result is a stabilized curvature approximation that accelerates convergence without compromising accuracy. This methodology is implemented in the Hessian-Engineered Relaxation Optimizer (HERO) package, which is openly available at <https://github.com/herodft-2026/beta>.

To maximize accessibility and facilitate seamless integration into established computational workflows, the HERO architecture is strictly designed within the framework of the Atomic Simulation Environment (ASE). By directly inheriting from the standard ASE Atoms and Optimizer classes, the package functions as a plug-and-play extension. This ensures a near-zero learning curve: any researcher familiar with the ASE ecosystem can deploy HERO immediately, utilizing it as a drop-in replacement for standard optimizers without the need to restructure their simulation scripts or data structures.

2 Methodology

We implemented this strategy within the ASE [21]—a Python library for working with atoms as a modular add-on, ensuring compatibility with widely used DFT codes. Benchmark tests on oxygen adsorption on Pt(111) surface demonstrate a substantial reduction in the number of ionic steps compared to standard BFGS. Additional tests on other adsorbates on Pt(111) and WO_x surfaces confirm the broad transferability of the method. By providing this scheme as an ASE extension, we aim to offer a practical and general solution for accelerating DFT structural optimization, with direct benefits for high-throughput studies in surface science and catalysis.

The flowchart in Fig. 1 delineates a robust and widely-used computational workflow for geometry optimization, a fundamental task in computational chemistry and materials science aimed at locating stable structures corresponding to local minima on the PES. The process is a synergistic combination of a high-accuracy quantum mechanical method (e.g., DFT) and an efficient numerical optimization algorithm (BFGS). The optimization process is iterative, beginning with an initial guess for the atomic structure, x_0 . Each iteration, indexed by k , proceeds as follows:

1. **Energy and Force Calculation:** The energy $E(x_k)$ and the forces on each atom are computed for the current structure x_k . The forces are equivalent to the negative gradient of the energy with respect to the atomic coordinates, $-\nabla E(x_k)$. As indicated, DFT is employed for this step, as it provides a good balance between accuracy and computational cost for a wide range of systems.
2. **Structure Update:** The atomic coordinates are updated to a new structure, x_{k+1} , in a direction that is expected to lower the system's energy. The flowchart specifies the use of the BFGS algorithm, a powerful quasi-Newton method. The update rule is given by:

$$x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla E(x_k),$$

where α_k is the step size and B_k is an approximation to the Hessian matrix (the matrix of second derivatives of the energy).

The inverse Hessian, B_k^{-1} , accounts for the local curvature of the PES, enabling a more intelligent and typically faster path to the minimum compared to simple gradient descent methods.

3. **Convergence Check:** After the update, the algorithm checks if a predefined convergence criterion has been met.

The management of the approximate Hessian matrix, B_k , is critical for the efficiency and stability of the BFGS algorithm. The ancillary information in Fig. 1 highlights a progressive strategy for this:

- a. **Block-Diagonalization (Spatial Topological Decoupling):** The most direct approach, shown in the flowchart, is to enforce a block-diagonal structure by zeroing the off-diagonal blocks ($B_{k,ij} \neq 0$ for $i \neq j$). This robustly enhances stability by eliminating potentially noisy coupling terms that can accumulate during the BFGS updates, which is especially useful in the early stages of optimization or for difficult systems.
- b. **Isotropic Background Stiffness Injection:** This block-diagonalization can be viewed as the baseline in a more advanced strategy. Instead of complete zeroing, a more nuanced approach involves

setting the off-diagonal blocks to small, controlled non-zero values. This acts as a form of dynamic structured Hessian preconditioning or preconditioning, providing a better balance between the stability of a diagonal Hessian and the faster convergence offered by a well-conditioned, dense Hessian. This allows for a more guided and efficient path to the energy minimum.

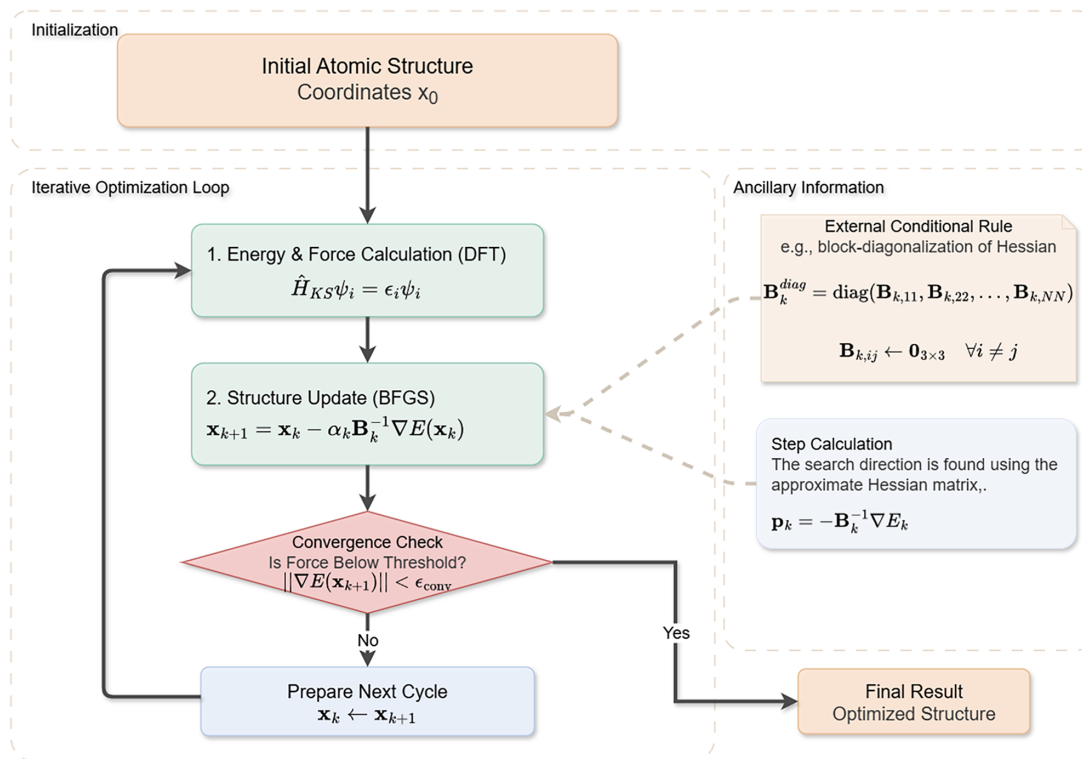


Figure 1: The iterative workflow for atomic structure optimization via the DFT-BFGS method. The process starts from an initial guess x_0 and converges to a final optimized structure. The flowchart also highlights a block-diagonalization scheme for the approximate Hessian matrix, used to enhance optimization stability.

As delineated in Schlegel’s review, the efficacy of geometry optimization—particularly within the framework of quasi-Newton schemes like BFGS—rests fundamentally on the validity of the local quadratic approximation of the potential energy surface [22]. However, challenges persist due to the numerical instability of Hessian updates in standard schemes, and the difficulty in handling the varying stiffness of molecular vibrational modes, which can degrade optimization efficiency [23]. In such regimes, standard update formulas tend to propagate gradient noise across all degrees of freedom [24], resulting in an Hessian that contains exaggerated off-diagonal elements and consequently overestimates the spurious inter-atomic couplings and local stiffness of the molecular system.

Addressing this limitation, our proposed strategy of uniform mean-field coupling (specifically the replacement of off-diagonal atomic blocks with a uniform damped constant) serves as a robust countermeasure to these instabilities. By strictly limiting the magnitude of these uncertain inter-atomic coupling terms in the iteration matrix, we effectively implement a dynamic spectral conditioning that acts as an intrinsic structural regularization. This operation aligns conceptually with the trust radius and rational function optimization (RFO) techniques discussed by Schlegel [22], as it prevents the optimizer from taking excessive steps driven by spurious “soft modes.” However, unlike extrinsic step-size constraints, our approach enforces

stability directly through the spatial decorrelation of the update operator, filtering out unreliable long-range noise while preserving the essential local topological information required for efficient convergence.

In essence, this represents a progression from a simple but stable “reset” of inter-atomic couplings to a more sophisticated isotropic background stiffness strategy that can accelerate convergence. On rugged potential energy surfaces, the strict decoupling of the ‘reset’ operation acts as a safety brake, preventing divergence but potentially stalling momentum in shallow valleys. The background stiffness strategy acts as a navigation aid: by enforcing a uniform, non-zero coupling, it smoothens the perceived local topology. This allows the optimizer to bypass minor corrugations that would otherwise trigger erratic Hessian updates, effectively guiding the system toward the local minimum with a trajectory that is both stable against anharmonic noise and dynamic enough to traverse flat regions of the PES efficiently.

3 Result

Mechanism of Hessian Reset (Fig. 2a–d): Panels (a) and (b) visualize the approximate Hessian matrix at intermediate steps of a geometry optimization. The significant non-zero off-diagonal elements (shown in red and blue) represent the learned couplings between different atomic coordinates. Over many iterations, particularly on a rugged energy landscape, these off-diagonal terms can accumulate numerical noise or represent non-local curvatures inaccurately, a phenomenon we term “Hessian pollution.” This can lead to inefficient or unstable optimization steps. The core of the HERO(0) method is a periodic “reset” of this matrix. As shown in the transition from (a) to (c) and (b) to (d), the algorithm identifies moments of stagnation or instability and sets the off-diagonal blocks to zero. This operation effectively purges the accumulated pollution, decoupling the coordinates and simplifying the local optimization problem to a set of independent blocks. Physically, this operation enforces a ‘locality constraint’ that the standard BFGS update lacks. In real materials, second-derivative couplings naturally decay with inter-atomic distance. However, numerical noise in the gradient approximation often smears across the entire matrix, creating artificial long-range correlations. By pruning these off-diagonal blocks, HERO explicitly aligns the mathematical model with the physical reality of short-range chemical interactions, effectively removing the ‘false memory’ of non-existent couplings. The resulting block-diagonal matrix in (c) and (d) provides a more robust, albeit less detailed, model of the local curvature, allowing the optimizer to escape from regions where the full Hessian approximation is problematic.

Performance and Convergence (Fig. 2e–f): The convergence curves demonstrate the practical benefit of this strategy. Panel (e) shows a challenging optimization scenario where the standard BFGS algorithm (blue line) suffers from significantly slower convergence, struggling to efficiently reduce the force below approximately 0.1 eV/\AA . This asymptotic stagnation is characteristic of an ill-conditioned Hessian, where the optimizer is misled by accumulated curvature errors into taking vanishingly small steps along shallow, non-productive search directions. In contrast, the HERO(0) algorithm (orange line) successfully navigates the complex energy landscape. The visible oscillations in the trajectory correspond to the algorithm actively resetting the Hessian to overcome optimization barriers. These transient spikes in force are not signs of numerical instability, but rather evidence of necessary ‘forgetting’. When the Hessian is reset, the optimizer discards the corrupted curvature history and momentarily reverts to a steeper descent direction (akin to a localized Gradient Descent restart). While this may cause a temporary increase in the projected force, it crucially breaks the inertia of the stagnation, allowing the system to re-orient towards the true local minimum. Crucially, this reset mechanism is triggered once the residual force falls into the activation window (below the threshold $F = 0.5 \text{ eV/\AA}$). Instead of the sluggish convergence observed in the standard method, this intervention actively purges Hessian pollution, enabling the optimizer to drive the forces down to the final convergence criterion (0.05 eV/\AA). Panel (f) exhibits a similar trend with a stricter activation threshold ($F =$

0.1 eV/Å), indicating that this parameter can be tuned to delay the intervention until a deeper convergence stage is reached.

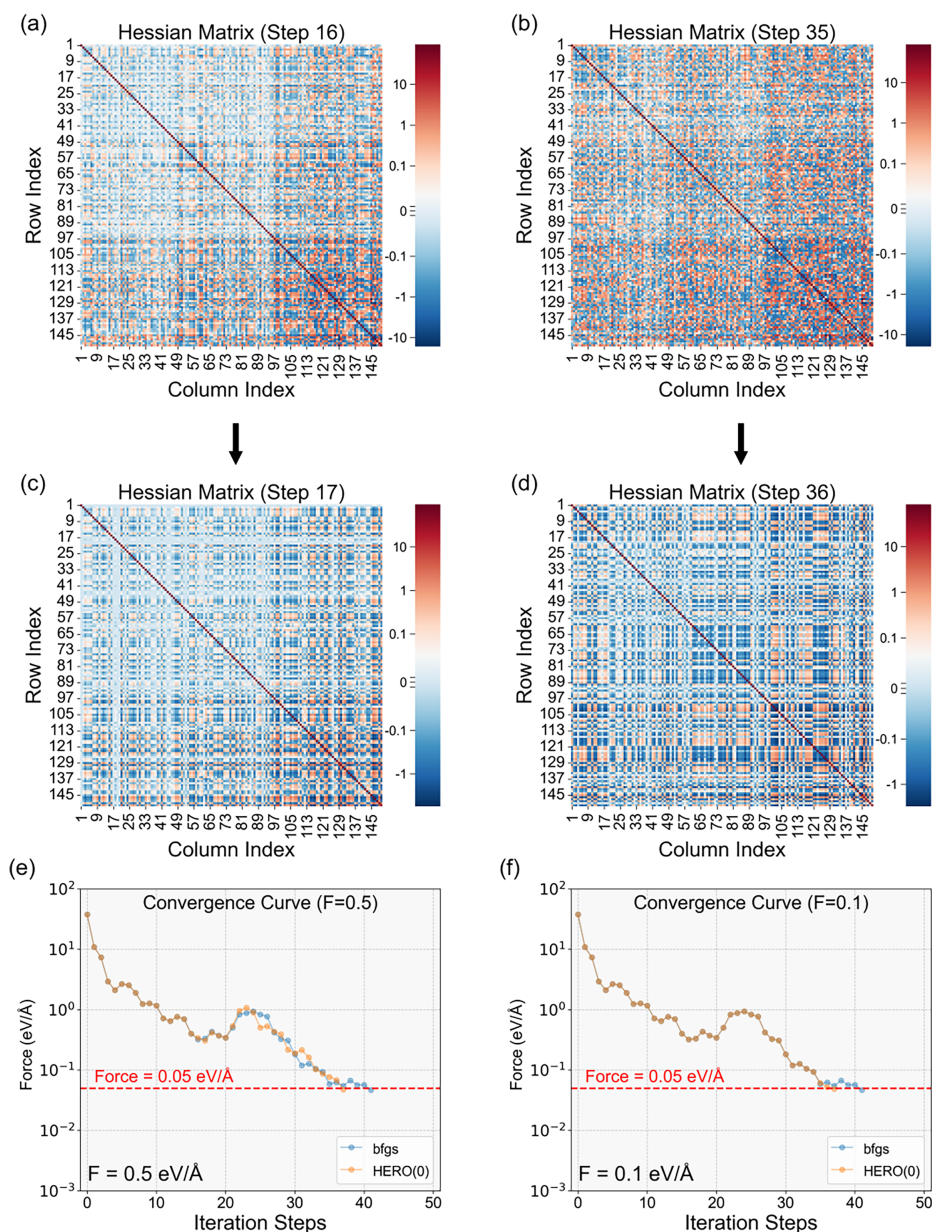


Figure 2: Illustration of the mechanism and performance of the HERO(0) optimization strategy, designed to enhance the robustness of quasi-Newton methods on complex potential energy surfaces. Panels (a–d) visualize the approximate Hessian matrix before and after the reset operation; specifically, the dense and “polluted” Hessians at optimization steps 16 and 35 (a,b) are transformed into clean, block-diagonal structures at steps 17 and 36 (c,d) by nullifying off-diagonal blocks. Correspondingly, panels (e,f) present convergence curves comparing HERO(0) (orange) against the standard BFGS method (blue) under two reset force thresholds ($F = 0.5 \text{ eV/Å}$ and $F = 0.1 \text{ eV/Å}$). The y-axis denotes the logarithm of the maximum atomic force, with the red dashed line indicating the final convergence criterion of 0.05 eV/Å .

The HERO(0) method provides a powerful enhancement to standard BFGS by dynamically simplifying the Hessian approximation. By periodically removing potentially erroneous coupling information, it prevents

the optimizer from getting trapped in difficult regions of the potential energy surface, proving particularly effective for systems where standard quasi-Newton methods struggle to converge.

The plots on the left (a, c, e, g) in Fig. 3 use a force threshold of $F = 0.5 \text{ eV/\AA}$ for Hessian resets, while those on the right (b, d, f, h) use a stricter threshold of $F = 0.1 \text{ eV/\AA}$. The “Hessian Engine” variants differ in the value added to the off-diagonal blocks during a reset (Engine0 adds 0, Engine4 adds 4, etc.). For each system, the insets show the initial and final atomic structures, confirming that all successfully converged methods yield the same final geometry. The x -axis represents the number of DFT force calls, a proxy for computational cost.

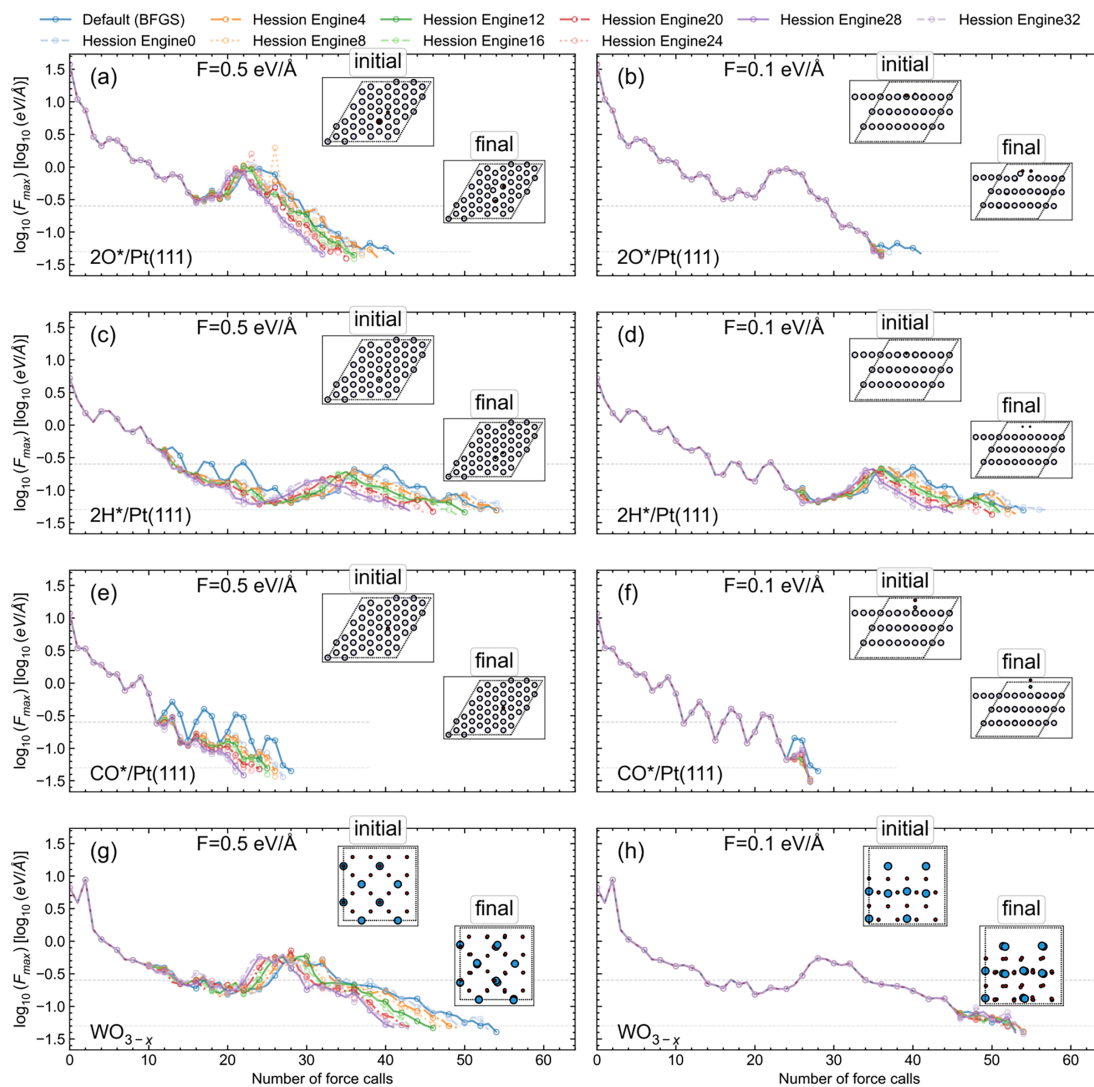


Figure 3: Comparative performance of various “Hessian Engine” optimization strategies against the default BFGS optimizer across four distinct chemical systems. The systems include adsorbates on a Pt(111) surface—(a,b) 2O^* , (c,d) 2H^* , and (e,f) CO^* —and (g,h) a defective WO_{3-x} structure. This selection is strategic: the atomic adsorbates (O^* , H^*) represent systems dominated by surface-hybridization modes, while the molecular adsorbate (CO^*) introduces high-frequency internal vibrations and rigid-body rotations that often create ill-conditioned Hessian eigenvalues. The defective bulk oxide (WO_{3-x}), conversely, tests the algorithm’s ability to handle long-range elastic relaxations and symmetry-breaking distortions typical of solid-state phase transitions.

Fig. 3 presents a comprehensive benchmark of our “Hessian Engine” optimization framework against a standard BFGS implementation. The tests span a range of materials, from surface-adsorbate systems to complex oxides, demonstrating the method’s broad applicability and robustness. Two key conclusions can be drawn from these results. These results demonstrate two key findings: first, that HERO accelerates convergence purely kinetically without introducing thermodynamic bias, consistently locating the identical final geometries; and second, that this degree of acceleration correlates strongly with the underlying complexity of the potential energy surface.

Crucially, the insets illustrate that all “Hessian Engine” variants converge to identical final geometries despite differing optimization trajectories, confirming that the method accelerates convergence without introducing artifacts or biasing the search. This confirms that the HERO modification is purely kinetic: it alters the trajectory through the potential energy landscape to bypass numerical traps, but it consistently converges to the identical local minimum without introducing thermodynamic bias to the final optimized geometry defined by the underlying DFT convergence criteria. This robustness is driven by an isotropic background stiffness injection strategy, wherein small positive values are added to off-diagonal blocks during truncation events. By mitigating the diagonal dominance typical of standard quasi-Newton initializations (which often enforces overly conservative step sizes), this approach improves Hessian conditioning and facilitates more aggressive optimization. Physically, a strictly diagonal Hessian (Engine0) implicitly assumes that atoms move independently, which is a poor approximation for condensed matter where collective motion is ubiquitous. By uniformly repopulating the off-diagonal blocks (e.g., Engine32), we effectively introduce a ‘mean-field’ background coupling. This heuristic restores a baseline level of inter-atomic connectivity, preventing the optimizer from underestimating the system’s collective response. This improves the conditioning of the approximate Hessian, effectively structuring the Hessian to enable more optimal and structurally consistent step sizes compared to the disconnected limit of Engine0. It is worth noting that the degree of acceleration provided by HERO correlates strongly with the complexity of the local PES. In systems where the local topology is already highly quadratic, standard BFGS performs near-optimally, and the HERO modifications naturally yield more marginal improvements. The true strength of the algorithm emerges in highly non-quadratic, rugged landscapes, where standard methods succumb to asymptotic stagnation. Indeed, increasing the off-diagonal addition (e.g., from Engine0 to Engine32) systematically reduces the required force calls, demonstrating that applying this structural preconditioning significantly expedites the optimization process compared to the default BFGS.

4 Implementation

Fig. 4 provides a schematic of the software architecture for our custom BFGS optimization engine, which is built upon the widely-used ASE Library. The architecture is designed for modularity, extensibility, and separation of concerns, which are crucial for developing and testing novel optimization strategies.

The system is organized into four primary layers connected by a central Computation Bus. The Application/Runtime layer, represented by the Optimization Runner, is responsible for initiating and managing the overall optimization task and injecting the necessary components. The Domain/Compute layer encapsulates the physics of the system, containing the ASE Atoms object and the Calculator interface that connects to external DFT codes. By strictly adhering to this standardized Calculator interface, the optimization core achieves code-agnostic versatility. It interacts with the underlying physics engine, supporting it be like VASP and Quantum ESPRESSO, or grid-based codes like GPAW including over 40 supported quantum chemistry and solid-state packages such as VASP, Quantum ESPRESSO, GPAW, CP2K, and Gaussian solely through abstract energy and force calls. Therefore, HERO is immediately adaptable to any DFT code supported by the broader ASE ecosystem without requiring user-side modifications. Finally, the Persistence/Logging layer,

with its State Manager, handles the saving and loading of the optimizer’s state, which is vital for restarting long calculations.

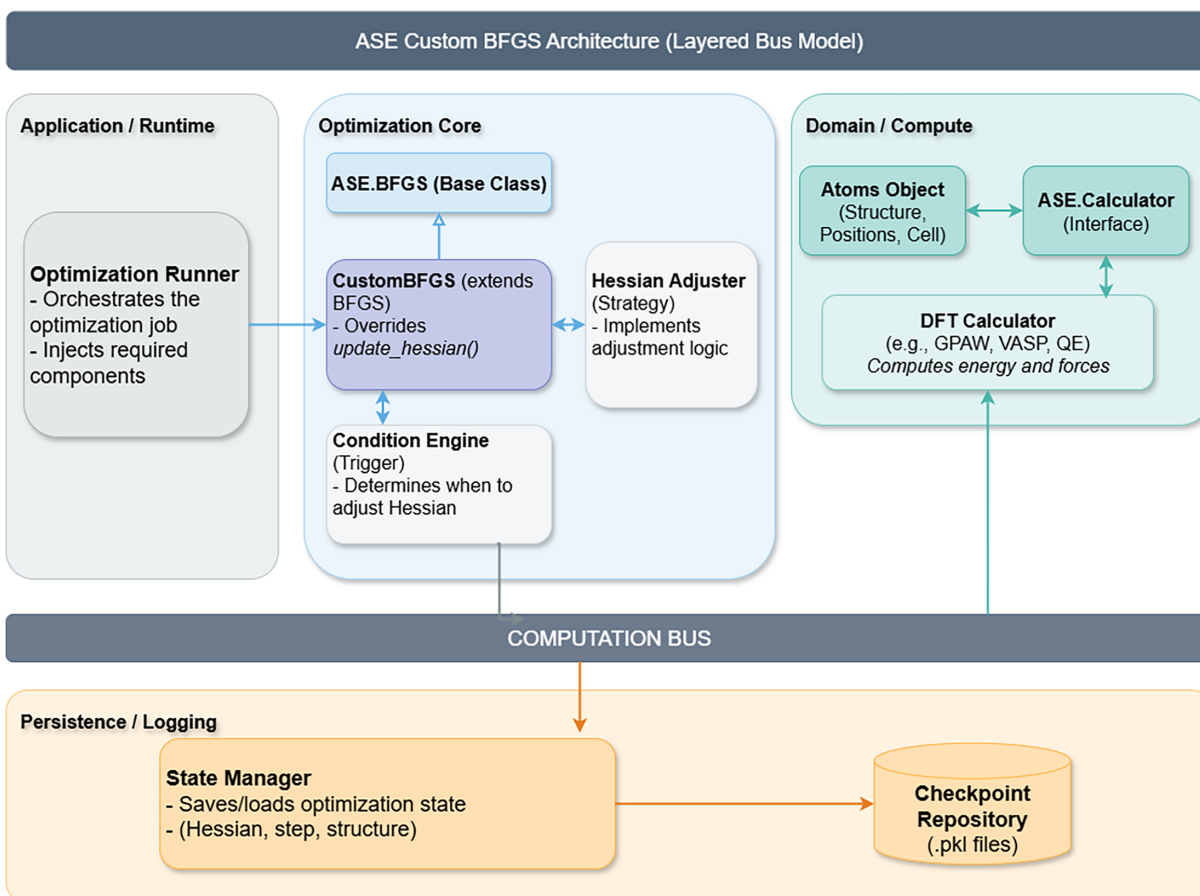


Figure 4: Software architecture of the custom BFGS optimizer, implemented within the atomic simulation environment (ASE) framework. The design follows a layered bus model, separating concerns into distinct logical domains: application/runtime, optimization core, domain/compute, and persistence/logging. The optimization core is modular, featuring a CustomBFGS class that extends ASE’s base BFGS implementation. It delegates the logic for “when” to adjust the Hessian to a condition engine (trigger) and “how” to adjust it to a Hessian adjuster (strategy). This modularity allows for flexible and extensible implementation of advanced optimization algorithms. All components communicate via a central computation bus, facilitating data flow and loose coupling between modules.

The key innovation lies within the modular Optimization Core. Instead of a monolithic optimizer, we have decomposed the logic using established software design patterns. The CustomBFGS class serves as the main driver, inheriting from ASE’s standard BFGS but overriding key methods to introduce our custom logic. The decision-making process for Hessian modification is decoupled into two distinct components: a Condition Engine (Trigger) that determines “if” and “when” a Hessian adjustment is needed based on criteria like force thresholds, and a Hessian Adjuster (Strategy) that implements the specific logic for “how” the Hessian should be modified, such as performing a block-diagonalization.

This decoupled design allows researchers to easily experiment with new optimization techniques by simply implementing new Condition Engine or Hessian Adjuster classes without altering the core optimization loop. This abstraction effectively ‘hot-swaps’ the standard BFGS logic with our Hessian-engineered approach. For the end-user, this implies seamless integration: HERO can be deployed in existing production scripts

with minimal changes—often requiring only the modification of a single import statement—while retaining full compatibility with the broader ASE ecosystem. This flexibility is fundamental to the rapid prototyping and validation of the advanced optimization algorithms presented in this work.

5 Conclusion

In this work, we have addressed the persistent challenge of “Hessian pollution” in geometry optimization, a phenomenon that significantly degrades the performance of quasi-Newton methods on non-quadratic potential energy surfaces. By identifying that high-order anharmonicities introduce spurious off-diagonal couplings into the approximate Hessian, we developed a physics-inspired modification to the standard BFGS algorithm that systematically filters this numerical noise.

The proposed strategy, implemented as the HERO package, relies on a conditional block-diagonalization scheme. Our benchmarks across diverse chemical systems (ranging from surface adsorbates on Pt(111) to defective bulk oxides) demonstrate that this approach effectively stabilizes the optimization trajectory. Crucially, we found that an isotropic background stiffness injection, which repopulates off-diagonal blocks with small positive constants rather than zeroing them completely, mitigates the excessive conservatism of diagonal dominance and further accelerates convergence.

Designed with a modular architecture within the ASE Library, this method offers a lightweight and transferable solution compatible with existing codes. Unlike black-box machine learning potentials that face transferability issues outside their training domain, HERO operates entirely on *ab initio* principles, requiring no prior data generation. By eliminating the need for complex parameter tuning or domain-specific training, HERO provides a robust tool for high-throughput computational workflows where autonomous stability is paramount.

Looking forward, the success of HERO confirms that algorithmic intervention at specific force thresholds is a critical “control knob” for stabilizing DFT calculations. Building on this, we envision a transition from the current uniform mean-field coupling to a more physically informed strategy. Future iterations could integrate physical priors such as bond-stiffness estimates derived from inter-atomic distances or coordination environments to construct a structured preconditioning matrix rather than a uniform diagonal one. Furthermore, there is significant potential in hybridizing this approach with advanced machine learning surrogate models (such as CHGNet or MACE) or empirical universal simple potentials (e.g., Morse or Lennard-Jones variants). By utilizing the analytical Hessians from these lightweight potentials as a baseline reference during the reset phase, we can inject robust physical intuition into the optimization process, effectively guiding the *ab initio* search with a “reasonable guess” that is computationally negligible yet topologically accurate.

Additionally, while our discussion has primarily framed “Hessian pollution” in the context of smooth, higher-order anharmonicities, it is crucial to note that real DFT potential energy surfaces are not strictly continuous. Phenomena such as spin-state transitions (e.g., in transition metal complexes) introduce genuine discontinuities and stochastic force jumps during ionic relaxation. From an algorithmic perspective, both smooth anharmonicity and true discontinuities represent severe violations of the local quadratic approximation. Standard BFGS struggles to recover from the non-reproducible gradient shocks caused by spin-state crossings. However, the block-diagonalization reset in HERO is intrinsically suited to address these discontinuous events. By periodically purging the iteration history, HERO effectively discards the corrupted, pre-transition curvature information, allowing the optimizer to rapidly re-adapt to the new PES branch. Thus, the HERO methodology generalizes beyond continuous anharmonicity to suppress pollution originating from non-differentiable PES features.

To facilitate community adoption and further development, the source code for this method is openly available at <https://github.com/herodft-2026/beta>.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by JSPS KAKENHI (No. JP25H01508), Gusu Laboratory of Materials (grant number Y2501), Suzhou MatSource Technology Co., Ltd., Beijing Natural Science Foundation (2262076), Natural Science Foundation of Hebei (E2025502039), and Fundamental Research Fund for the Central Universities (2025JC008 and 2025MS131).

Author Contributions: Mingzhe Li: Methodology, Software, Formal analysis, Writing—original draft. Piao Ma: Methodology, Software, Formal analysis. Limin Li: Methodology, Software, Formal analysis. Weijie Yang: Methodology, Supervision. Hao Li: Conceptualization, Investigation, Formal analysis, Writing—original draft. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: All test data is available on GitHub.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jain A, Shin Y, Persson KA. Computational predictions of energy materials using density functional theory. *Nat Rev Mater*. 2016;1(1):15004. doi:10.1038/natrevmats.2015.4.
2. Car R, Parrinello M. Unified approach for molecular dynamics and density-functional theory. *Phys Rev Lett*. 1985;55(22):2471–4. doi:10.1103/physrevlett.55.2471.
3. Payne MC, Teter MP, Allan DC, Arias TA, Joannopoulos JD. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev Mod Phys*. 1992;64(4):1045–97. doi:10.1103/revmodphys.64.1045.
4. Head JD, Zerner MC. A Broyden-Fletcher–Goldfarb-Shanno optimization procedure for molecular geometries. *Chem Phys Lett*. 1985;122(3):264–70. doi:10.1016/0009-2614(85)80574-1.
5. Pfrommer BG, Côté M, Louie SG, Cohen ML. Relaxation of crystals with the quasi-Newton method. *J Comput Phys*. 1997;131(1):233–40. doi:10.1006/jcph.1996.5612.
6. Gill PE, Leonard MW. Reduced-hessian quasi-Newton methods for unconstrained optimization. *SIAM J Optim*. 2001;12(1):209–37. doi:10.1137/s1052623400307950.
7. Instituto de Matemática Pura e Aplicada. White paper: complex energy landscapes. Rio de Janeiro, Brazil: IMPA; 2018.
8. Oren SS. Quasi-Newton algorithms: approaches and motivations. In: 1973 IEEE Conference on Decision and Control Including the 12th Symposium on Adaptive Processes; 1973 Dec 5–7; San Diego, CA, USA. p. 422–7. doi:10.1109/CDC.1973.269202.
9. Mones L, Ortner C, Csányi G. Preconditioners for the geometry optimisation and saddle point search of molecular systems. *Sci Rep*. 2018;8(1):13991. doi:10.1038/s41598-018-32105-x.
10. Vandermause J, Torrisi SB, Batzner S, Xie Y, Sun L, Kolpak AM, et al. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput Mater*. 2020;6(1):20. doi:10.1038/s41524-020-0283-z.
11. Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel CJ, et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat Mach Intell*. 2023;5(9):1031–41. doi:10.1038/s42256-023-00716-3.
12. Zhang D, Chen Y, Liu C, Liu Y, Xin H, Peng J, et al. Accelerating catalyst materials discovery with large artificial intelligence models. *Angew Chem Int Ed*. 2026;65(16):e26150. doi:10.1002/anie.202526150.

13. Kovács DP, Batatia I, Arany ES, Csányi G. Evaluation of the MACE force field architecture: from medicinal chemistry to materials science. *J Chem Phys.* 2023;159(4):044118. doi:10.1063/5.0155322.
14. Fu C, Lin Y, Krueger Z, Yu W, Qian X, Yoon BJ, et al. A benchmark for quantum chemistry relaxations via machine learning interatomic potentials. arXiv:2506.23008. 2025.
15. Zhao C, Ma Z, Fan D, Hu S, Wang L, Jia W, et al. Integrating machine learning interatomic potentials with batched optimization for crystal structure prediction. 2025. doi:10.26434/chemrxiv-2025-fqs7n.
16. Li Y, Zhang X, Liu M, Shen L. A critical review of machine learning interatomic potentials and Hamiltonian. *J Mater Inf.* 2025;5(4):43. doi:10.20517/jmi.2025.17.
17. Zhang D, Jia X, Wang Y, Liu H, Wang Q, Jang SH, et al. Digital materials ecosystem: from databases to AI agents for autonomous discovery. *Chem Sci.* 2026;17(12):5782–804.
18. Bitzek E, Koskinen P, Gähler F, Moseler M, Gumbusch P. Structural relaxation made simple. *Phys Rev Lett.* 2006;97(17):170201. doi:10.1103/physrevlett.97.170201.
19. Wang LP, Song C. Geometry optimization made simple with translation and rotation coordinates. *J Chem Phys.* 2016;144(21):214108. doi:10.1063/1.4952956.
20. Xie Y, Vandermause J, Sun L, Cepellotti A, Kozinsky B. Bayesian force fields from active learning for simulation of inter-dimensional transformation of stanene. *npj Comput Mater.* 2021;7(1):40. doi:10.1038/s41524-021-00510-y.
21. Hjorth Larsen A, Jørgen Mortensen J, Blomqvist J, Castelli IE, Christensen R, Dułak M, et al. The atomic simulation environment—a Python library for working with atoms. *J Phys Condens Matter.* 2017;29(27):273002. doi:10.1088/1361-648x/aa680e.
22. Schlegel HB. Geometry optimization. *WIREs Comput Mol Sci.* 2011;1(5):790–809. doi:10.1002/wcms.34.
23. Birkholz AB, Schlegel HB. Exploration of some refinements to geometry optimization methods. *Theor Chem Acc.* 2016;135(4):84. doi:10.1007/s00214-016-1847-3.
24. Shi HM, Xie Y, Byrd R, Nocedal J. A noise-tolerant quasi-Newton algorithm for unconstrained optimization. *SIAM J Optim.* 2022;32(1):29–55. doi:10.1137/20m1373190.