# Simulating Atoms by Sampling Generative Models: Towards Non-Equilibrium Processes

Andreas Burger
andreas.burger@mail.utoronto.ca
Department of Computer Science
University of Toronto
Canada

June 17, 2025

**Abstract**

Molecular simulations are a useful tool in chemistry to predict outcomes where experiments are difficult to perform, and to gain insights into phenomena that cannot be measured in the lab. Direct simulation via molecular dynamics or Monte Carlo is limited by long autocorrelation times and poor exploration of high-barrier regions. Molecular dynamics and Monte Carlo can be interpreted as samplers from the Boltzmann distribution, where the central problem becomes one of efficiently generating independent samples.

Generative models offer a promising alternative approach to traditional sampling methods. Once trained, they generate uncorrelated, Boltzmann-distributed samples efficiently, bypassing the slow dynamics of conventional simulations. A key challenge is the sampling of rare events, that govern the dynamics of molecular systems. Transition states in particular act as bottlenecks between metastable states and determine the rates of chemical reactions, drug binding, or protein folding. Observables sensitive to these rare configurations are often underestimated or missed entirely in standard simulations.

This paper reviews central developments in sampling equilibrium distributions and transitions with generative models. We conclude that while generative modeling has made significant progress in sampling equilibrium distributions, existing methods for rare transitions fail to be practically useful due to limited training data or computationally expensive optimization requirements for each individual transition. To address this gap, we propose a diffusion sampler that can sample a diverse set of unknown transitions by training on the limited data that exists, as well as training directly on energy landscapes where experimental data is scarce.

We conclude by providing a unified perspective that argues for a natural progression in the field of generative sampling of molecules: from single structures, to equilibrium ensembles, to individual transition states, and ultimately to non-equilibrium path ensembles.

## 1 Introduction - Simulation as Generative Modeling

Molecular simulations serve as essential tools in chemistry to predict outcomes where experiments are difficult to perform and to gain insights into phenomena that cannot be measured experimentally.

The fundamental challenge lies in the computational expense of these simulations. Traditional molecular dynamics simulations require integration of equations of motion using timesteps on the order of femtoseconds to accurately capture the fastest events, like hydrogen bond vibrations. However, many phenomena of scientific and practical interest occur on dramatically longer timescales. Protein folding events unfold over microseconds to milliseconds, drug binding processes span nanoseconds to microseconds, and catalytic reactions may require picoseconds to nanoseconds. This temporal disparity creates a computational bottleneck where approximately $10^9$ molecular dynamics simulation steps are needed to observe a single protein folding event, rendering such calculations prohibitively expensive.

The connection between atom-level simulations and macro-level laboratory observations requires computing expectation values over ensemble distributions. Predicting experimental measurements necessitates averaging over numerous microscopic simulation states, which demands observing phenomena many

times to achieve statistical convergence. This sampling requirement compounds the computational challenge, as accurate expectation values require extensive statistical sampling across the relevant phase space.

Transition states represent a particularly critical aspect of this sampling problem. These configurations determine key processes in drug discovery, including molecular docking, conformational changes, and chemical reactions. From a statistical mechanics perspective, transitions constitute rare events with low probability that can nonetheless dominate expectation values due to their outsized influence on observable properties Jarzynski (2006). The infrequent occurrence of these events makes them computationally expensive to capture through conventional molecular dynamics approaches, yet understanding these transition states provides essential mechanistic insights that guide rational design of catalysts, pharmaceuticals, and protein mutations.

Recent advances in generative modeling have opened new avenues for addressing these computational challenges by reframing simulation as a sampling problem. Rather than following deterministic trajectories through phase space, generative models can be trained to produce samples directly from the equilibrium Boltzmann distribution, potentially bypassing the slow dynamics that limit traditional approaches. This perspective transforms the fundamental computational task from integrating equations of motion to efficiently generating independent samples from target probability distributions.

This review examines the current state and future prospects of generative modeling approaches to molecular simulation. The first section discusses methods that address the timescale problem by efficiently generating independent samples from the Boltzmann distribution for equilibrium properties. The second section focuses on specialized techniques for sampling transition states and transition pathways, which represent the frontier of current research efforts. Through this analysis, we aim to establish a comprehensive framework for understanding how generative modeling can revolutionize computational chemistry and molecular simulation.

## 2 Sampling the Boltzmann distribution

The fundamental purpose of molecular simulation is to compute expectation values. An analogy would be the simulation of a self driving car. We might simulate many hours of driving to answer the question: "How often (and when) does the car crash?", an expectation value. This framework captures the nature of simulation objectives: experimental measurement, thermodynamic property, and material characteristic that we seek to predict through simulation correspond to expectation values. Internal energy represents the expectation of the potential energy function, heat capacity emerges from energy fluctuations, structural correlation functions arise from position-dependent observables, and phase transition order parameters correspond to specific configuration-dependent functions Frenkel & Smit (2002). The computational challenge of molecular simulation therefore reduces to the efficient estimation of these expectation values. Simulation can thus be seen as a sampling problem from high-dimensional probability distributions of atomic configurations.

**The Canonical Ensemble and Equilibrium Distribution** The first question is what is the probability distributions of atomic configurations that we should sample from. The most common model is to assume that the system has relaxed to thermal equilibrium, corresponding to infinite time. The equilibrium distribution $\mu(x)$ gives the probability of finding a system in state $x$. This distribution proves particularly convenient for computational purposes because macroscopic quantities become time-independent, allowing meaningful comparison with experimental measurements.

We can describe this equilibrium as a system of $N$ atoms, where $x_n = (x_{n,1}, x_{n,2}, x_{n,3})$ represent the position coordinates of the $n$-th particle in 3D space, and $x = (x_1, \ldots, x_N)$ denote the complete configuration of the system. The system maintains contact with a heat bath at fixed temperature $T$, allowing energy exchange while keeping particle number $N$ and volume constant. We assume that the atoms are subject to an energy $E$, which we can compute with methods like DFT or machine learning force fields Aldossary et al. (2024). The equilibrium distribution of such a system, known as the canonical or NVT ensemble, is given by the Boltzmann distribution:

$$p(x) = \mu(x) = \frac{1}{Z_V} \exp[-\beta E], \tag{1}$$

where $Z_V = \int \mathrm{d}x \exp[-\beta E]$ represents the canonical partition function and $\beta = 1/k_B T$ denotes the inverse temperature, with $k_B$ being Boltzmann's constant.

Once $\mu(x)$ is known, any observable $A(x)$ has expectation value

$$\mathbb{E}_{p(x)}[A(x)] = \langle A \rangle = \int A(x)\mu(x)dx \tag{2}$$

These averages correspond directly to thermodynamic quantities such as internal energy, magnetization, and structural properties that can be measured experimentally. The partition function connects directly to the Helmholtz free energy $F$

$$F = -\beta^{-1} \ln Z_V. \tag{3}$$

In practice, we are interesting in estimating two primary quantities: expectations under the Boltzmann distribution $\mathbb{E}_{p(x)}[A(x)]$ of physical observables $A$, and the partition function $Z_V$ itself, which enables calculation of the free energy $F$. These problems present significant computational challenges due to the high dimensionality of the configuration space $d = 3N$.
The following discussion focuses specifically on generative methods for sampling the Boltzmann distribution in the canonical ensemble, which provides the foundation for most molecular simulation applications.

## 2.1 Boltzmann Generators: Sampling Equilibrium States Of Many-Body Systems With Deep Learning

**Method and impact** Boltzmann generators Noé et al. (2019) were arguably the first approach to sample from a generative model instead of long molecular dynamics trajectories. Once trained, sampling from the neural network allows to compute expectation values like free energy differences. Different to the sequential steps of MD, samples from the model are statistically independent.
They showed a discrete normalizing flow $F$, the Boltzmann generator, could be trained on a mixture of data (training-by-example) using maximum likelihood estimation as well as directly on an energy function $E$ (training-by-energy) like a force field or DFT Aldossary et al. (2024).

$$L_{ML} = \mathbb{E}_x \left[ \frac{1}{2} ||F^{-1}(x)|| - \log |\det J^{-1}(x)| \right], \; J^{-1}(x) = \begin{bmatrix} \frac{\partial F^{-1}(x)}{\partial x_1} & \dots & \frac{\partial F^{-1}(x)}{\partial x_n} \end{bmatrix} \tag{4}$$

$$L_{KL} = \mathbb{E}_z \left[ E(F(z)) - \log |\det J(z)| \right], \; J(z) = \begin{bmatrix} \frac{\partial F(z)}{\partial z_1} & \dots & \frac{\partial F(z)}{\partial z_n} \end{bmatrix} \tag{5}$$

Where the latent variable $z$ is sampled from Gaussian noise $z \sim \mathcal{N}(0, 1)$.
If the initial data is biased, as usually the case with finite-length MD, one can reweight the samples. First, the Boltzmann Generator can be used to generate unbiased samples by sampling $x \sim \tilde{p}(x)$ with the normalizing flow. then computing corresponding importance weights $w(x) = \mu(x)/\tilde{p}(x)$ for each sample. These allow to reweight generated samples to the target Boltzmann distribution $\mu(x)$. It is possible to estimate observables of interest (asymptotically unbiased) using the weights $w(x)$ with importance sampling via

$$\langle O \rangle_\mu = \frac{\mathbb{E}_{x\sim\tilde{p}(x)}[w(x)O(x)]}{\mathbb{E}_{x\sim\tilde{p}(x)}[w(x)]}. \tag{6}$$

This is possible because the normalizing flow provides exact likelihoods for the generated samples.
To find transitions between metastable states (energy minima) they showed one could use Monte Carlo in latent space for exploration and latent interpolation to estimate transition states.

**Limitations** There are multiple caveats to Boltzmann Generators. The model needs to be initialized with data covering each metastable state, which requires running many MD or MC chains for data collection. This is because training-by-energy with the KL-divergence leads to mode collapse Vargas et al. (2022); Richter & Berner (2024).
To by using sampling methods such as Metropolis Monte Carlo in the latent space, Boltzmann generators can discover new states and gradually explore state space. used to explore in latent space while training the network. problem is that computing acceptance probability of a step requires going to configuration space, making generating proposals expensive.
To get accurate free-energies, the reweighting step is necessary. While this reduces the requirements on the data to be unbiased, the further away from the true Boltzmann distribution the more samples are required for importance sampling.

The paper also argued that simple linear interpolations between points in latent space can correspond to low-energy transition pathways. Only shown experimentally on two dimensional toy Mueller. Known that geometric techniques such as geodesic interpolation and optimal transport better align with the intrinsic manifold structure of the data.
Lastly, the method relies on exact-likelihood discrete normalizing flows. These models are known to be poorly scalable, due to the need to backpropagate through generation chain as opposed to a single noising-denoising step like in diffusion, dubbed simulation-based training. Normalizing flows also put severe restrictions on the architecture, with cannot easily be integrated with graph networks that are state of the art for molecular data.

## 2.2 Equivariant Flow Matching

**Method and impact** Equivariant flow matching Klein et al. (2023) followed up on Boltzmann generators by making it considerably more scalable. Scalability is a key consideration, as retraining a new model on each system quickly negates most advantages of ML-based sampling Tan et al. (2025).
The first improvement was to replace the discrete normalizing flow with a continuous normalizing flow. This CNF can now be trained simulation free using the flow matching loss Lipman et al. (2022) instead of maximum likelihood, i.e. without integrating the vector field and evaluating the Jacobian, making the training significantly faster. The use of a CNF also removes the need for the architecture to be invertible, which allows for graph network architectures that can transfer across molecules with different number of atoms.
The second improvement was to introduce SE(3) equivariance to the architecture using features based on spherical harmonics Thomas et al. (2018); Satorras et al. (2021). Known to improve generalization when data is limited Klein & Noe (2024); Batzner et al. (2022).
In addition, the generation paths can be improved with optimal transport Tong et al. (2023), which allows for faster inference.

**Limitations** Switching from discrete to continuous normalizing flows, while faster, has the downside that one looses access to the exact likelihoods of samples.
As a consequence, reweighting becomes more expensive, as it requires $\log p(x)$ via $\frac{d \log p(z(t))}{dt} = -\operatorname{Tr}\big[\partial_z F(z(t),\, t)\big]$. Evaluating that trace directly is $O(d^2)$, so one usually uses Hutchinson's estimator Skreta et al. (2024) $\operatorname{Tr}\big[\partial_z f\big] = \mathbb{E}_v\big[v^T\, \partial_z F\, v\big]$, using a random vector $v$ with $\mathbb{E}[vv^T] = I$, which costs $O(d)$ per one Jacobian-vector product.
While it allows for faster inference, the approximation of optimal transport requires matching the atoms using the Hungarian algorithm, which has a computational complexity of $O(N^3)$.
Most importantly, using a simulation-free objective for flows[1] no longer allows for training by energy. This means data covering all modes has to be precollected, which puts the burden of exploration on a classical sampling method.

## 2.3 Iterated Denoising Energy Matching for Sampling from Boltzmann Densities

**Method and impact** DEM Akhound-Sadegh et al. (2024) kept the equivariant architecture similar to EFM Satorras et al. (2021), but proposes a simulation-free training objective that uses solely the energy function and its gradient without requiring data samples.
DEM trains a diffusion-based sampler, parametrizing the score $s_\theta$, by alternating between two stages. In the outer loop DEM generates samples from the diffusion score model and saves the samples in a replay buffer. In the inner loop the score model is optimized based on the generated samples. The true score $\nabla \log p_t(x_t)$ is approximated using $k$ Monte Carlo samples

$$\nabla \log p_t(x_t) \approx S_K(x_t, t) = \frac{\frac{1}{K}\sum_i \nabla \exp(-E(x_{0|t}^{(i)}))}{\frac{1}{K}\sum_i \exp(-E(x_{0|t}^{(i)}))} = \nabla_{x_t} \log \sum_i \exp(-E(x_{0|t}^{(i)}))$$

[1] Here we use "flow models" as an umbrella term for the closely related family of generative methods around continuous normalizing flows, flow matching, denoising diffusion models, score-based models, etc.

or equivalently

$$S_K(x_t, t) = -\sum_i w_i \nabla E(x_{0|t}^{(i)}), \quad w_i := \frac{\exp(-E(x_{0|t}^{(i)}))}{\sum_j \exp(-E(x_{0|t}^{(j)}))} \propto p_0(x_{0|t}^{(i)}).$$

Where $x_{0|t}^{(1)}, \ldots, x_{0|t}^{(K)} \sim \mathcal{N}(x_t, \sigma_t^2 I)$. Intuitively, DEM effectively bootstraps a dataset by using the replay buffer, while generating samples from the imperfect model acts as an exploration mechanism.

**Limitations** The importance weighting loss in Eq. 7 has two major drawbacks. The first is that the estimator becomes high variances further away the generated samples are from the true distribution, which becomes worse in higher dimensions. To reduce the variance one has to compensate with large batch sizes. Atomistic systems can be especially challenging, since forces (same as the score $F = \nabla_x E$ diverge to infinity as atoms come closer and tend towards zero as atoms pull apart.
The second downside of the loss is that it requires an overwhelming number of energy evaluations, one for each Monte Carlo sample $k$ for each sample $x$ at each gradient descent step. Even thought DEM does not require any data samples, the cost of optimization, in terms of energy function calls, ends up being larger than precollecting data and training a regular diffusion model. He et al. (2025). This eventually limited the results to synthetic toy problems instead of physical molecules.

## 2.4 Adjoint Sampling: Highly Scalable Diffusion Samplers via Adjoint Matching

**Method and impact** Adjoint Sampling Havens et al. (2025) introduces key improvements to be able to scale to sample conformers in small molecules up to 30 heavy atoms. Adjoint Sampling optimizes an equivalent stochastic optimal control problem

$$\min_u \mathbb{E}_{X \sim p^u} \left[ \int_0^1 \frac{1}{2} \|u(X_t, t)\|^2 \, dt + g(X_1) \right] \tag{7}$$

$$\text{s.t.} \quad dX_t = \sigma(t) u(X_t, t) \, dt + \sigma(t) \, dB_t, \quad X_0 = 0 \tag{8}$$

where $\sigma : [0, 1] \to \mathbb{R}$ is a scalar noise function, $u : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ is a learnable control parametrized by a neural network, and $(B_t)_t$ is a $d$-dimensional Brownian motion.
Adjoint Sampling keeps the two-stage approach of DEM with a replay buffer, but proposes the new Reciprocal Adjoint Matching (RAM) objective Domingo-Enrich et al. (2024)

$$\mathcal{L}_{\text{RAM}}(u) = \int_0^1 \frac{1}{\sigma(t)^2} \mathbb{E}_{X_t \sim p_{t|1}^{\text{base}}, X_1 \sim p_1^{\tilde{u}}} \left[ \frac{1}{2} \|u(X_t, t) + \sigma(t) \nabla g(X_1)\|^2 \right] dt \tag{9}$$

$$g(x) = \log p_1^{\text{base}}(x) + \frac{1}{\tau} E \tag{10}$$

The objective significantly reduces the number of energy evaluations, since the energy function only needs to be called once per sample in the outer loop, but no longer in the inner loop at every SGD step. Practically the loss is lower variance, and better suited for atomistic systems with exploding and vanishing forces. Interestingly, the objective only depends on the force (score) of the energy, but not the energy itself.
If some data $p^{\text{data}}(X_1)$ is available, the NN control $u_\theta$ can be pretrained using Bridge Matching Shi et al. (2023).

$$\mathcal{L}_{\text{BM}}(\theta) = \int_0^1 \mathbb{E}_{p_{t|1}^{\text{base}}(X_t|X_1) p^{\text{data}}(X_1)} \left\| \frac{\nu_{1|t}}{\sigma(t)} u_\theta(X_t, t) - (X_1 - X_t) \right\|^2 dt, \quad \nu_{1|t} = \int_t^1 \sigma(s)^2 \, ds \tag{11}$$

which can reduce the cost of training significantly.

**Limitations** Adjoint Sampling test their method on sampling conformer states of medium-sized molecules. To preserve the bonds of the molecules they need to introduce an additional force that constraints bonds around length typical for involved atom types. This does not guarantee the correct connectivity and samples which change the molecular graph need to be removed in postprocessing.

Similar to DEM, Adjoint Sampling relies on the outer loop sample generation to explore the potential energy surface. In practice, large batch sizes are required to achieve mode coverage and not collapse to few minima.
Lastly, there currently is no mechanism to sample the modes of the learned distribution during inference. The reported evaluation relies on ground-truth samples, as they compute the pairwise root mean squared distance between all generated and ground-truth samples. To assign generated samples to conformations the sampled conformations are filtered based on distance to the true conformations. Even when producing and filtering a large number of samples, Adjoint Samplings improves recall but yields worse precision than RDKit on generalization on the Geom-Drugs dataset.

### 2.5 Alternative Statistical Ensembles

While the Boltzmann distribution represents the most commonly encountered model in molecular simulation, other stationary distributions exist and serve important roles in computational chemistry and materials science. The isobaric-isothermal (NPT) ensemble maintains constant pressure rather than volume, making it particularly relevant for chemistry applications where experimental conditions typically involve atmospheric pressure. This ensemble enables estimation of Gibbs free energies, analogous to how the canonical NVT ensemble facilitates computation of Helmholtz free energies. Previous work has addressed the based sampling of the NPT ensemble with Boltzmann Generators by parametrizing certain types of volume changes Wirnsberger et al. (2023); van Leeuwen et al. (2023). The grand canonical ensemble ($\mu$VT) presents even greater difficulty due to fluctuations in particle number, and, to the best of our knowledge, remains an open problem.

## 3 Generative Modeling of Transition States and Paths

The previous section focused on method sampling Boltzmann distributions. The major limitation is that they fail to describe processes. Processes are often characterized by rare events, which have a low probability under the equilibrium distribution. While being rarely sampled, rare events can dominate expectation values related to processes Jarzynski (2006). Chemical reactions, conformation changes of molecules, drug binding, and crystallization are examples of rare events.
Transition-state theory (TST) has been the foundation for the development of a large number computational tools for studying barrier-crossing events. What is most attractive about TST is its simplicity. It states that transitions like reactions are determined by the free energies of two types of stationary points: the minima and the transition state. The transition state acts as a bottleneck that has to be overcome, which is a saddle point on the potential energy surface. One can then approximate the reaction rate from the transition-state using Eyring equation Eyring (1935)

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}} = \frac{k_B T}{h} e^{\frac{\Delta S^\ddagger}{R}} e^{-\frac{\Delta H^\ddagger}{RT}} \tag{12}$$

where $k_B$ is Boltzmann's constant, $h$ is Planck's constant, $\Delta H^\ddagger$ is the activation enthalpy, $\Delta S^\ddagger$ is the activation entropy, and $\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger$. In many cases, one can also define the most probable transition path for the reaction, which for overdamped systems is simply the minimum energy path (MEP).
While a useful approximation for a wide range of systems, TST is limited to situations in which the potential energy surface is rather smooth and there is no recrossing through the transition-state region. If the crossing is diffusive (recrossings), then TST overestimates the reaction rate. For a system with a rugged potential energy landscape, or when entropic (i.e., volume) effects matter (as they typically do in high dimensions), the saddle points do not necessarily play the role of transition states.
Another popular approach is to consider the free energy surface instead of the potential energy surface, because the free energy surface is typically much more smooth. However, to define the free energy, one must have a set of collective variables to begin with, and therefore the concept of reaction coordinates has been introduced. Intuitively, reaction coordinates should be something that can be used to parameterize the reaction paths. Often one has to guess what the reaction coordinates are based on intuition and can often be hard to guess.
A more complete viewpoint is provided by the transition-path sampling (TPS) technique Pratt (1986); E & Vanden-Eijnden (2010). Instead of focusing on the transition states, TPS focuses on the ensemble

of transition path by providing is a way of Monte Carlo sampling the transition-path ensemble. By considering the full distribution of paths, accurate reaction rates can be computed.
More generally, the equilibrium model breaks down if detailed balance (reversability) is broken. Examples include slowly relaxing systems that may never reach true equilibrium on any practical timescale, significant fluctuations that break ensemble equivalence, or driving with time-dependent external fields.
Non-equilibrium thermodynamics fundamentally requires theories of path ensembles. For example, free energy differences $\Delta F$ are in general given by the Jarzynski equality Jarzynski (1997) $\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}$. Within this framework, work becomes a functional $W[\Gamma]$ over trajectories $\Gamma$ connecting states A and B, and the ensemble average takes the form $\langle e^{-\beta W} \rangle = \int \mathcal{D}\Gamma P[\Gamma] e^{-\beta W[\Gamma]}$. Crooks fluctuation theorem $P[\Gamma]/\tilde{P}[\tilde{\Gamma}] = e^{\beta(W[\Gamma]-\Delta F)}$ generalizes Jarzynski's equality to relate forward/reverse path probabilities in path ensembles Crooks (1999).

**Traditional Methods for Rare Events** There is a rich literature on classical methods for transition state and transition path sampling that are widely used, despite their limitations.
Common double-ended approaches requires knowledge of the minima (i.e. geometries of reactants and products), to then perform costly optimization for each pair of stable states Henkelman et al. (2000); Peters et al. (2004); Zimmerman (2015).
Single-ended approaches such as artificial force-induced reaction (AFIR) Maeda et al. (2016), anharmonic downward distortion following (ADDF), and single-ended growing string methods Zimmerman (2015) start from reactant geometries alone but scale combinatorially with system size, making large molecular systems expensive to study. They also rely on hand-crafted heuristics, which can fail for less common systems.
Enhanced sampling methods can explore the potential energy surface but have their own problems. Umbrella sampling and metadynamics Laio & Parrinello (2002); Hénin et al. (2022) apply biasing potentials along collective variables to speed up sampling of rare events. Choosing appropriate collective variables requires chemical insight and often involves trial-and-error optimization. Poor collective variable choices can result in incomplete sampling or incorrect results.
Shooting techniques sample dynamical trajectories using Metropolis-Hastings criteria to connect different stable states. While these methods avoid predefined collective variables, they suffer from slow sampling due to high rejection rates when proposing new trajectory segments. The computational cost of generating long molecular dynamics trajectories for each proposal makes these methods inefficient.
These limitations of classical approaches motivate the development of generative modeling techniques. The goal is to reduce both computational cost and the chemical intuition required for transition state exploration. Generative methods aim to create more automated and efficient approaches to rare event sampling that can discover new transitions without extensive prior knowledge of the system's energy landscape.

## 3.1 Diffusion Models For Transition States: OA-ReactDiff, TSDiff

**Method and impact** OA-ReactDiff Duan et al. (2023) and TSDiff Kim et al. (2024) are two concurrent works that address the problem of finding transition state configurations by learning a diffusion model. OA-ReactDiff predicts the highest energy image of the DFT-based climbing image NEB Henkelman et al. (2000), conditioned on the reactant and product geometries. During training OA-ReactDiff jointly generates reactant, TS structure, and product concatenated together. During inference the reactant and/or product can be fixed and the TS structure is generated via inpainting. TSDiff instead only generates the transition state conditioned on the smiles representation of the reactant.

**Limitations** Both methods are limited by requiring at least one minima (i.e. either reactant or product) of the reaction beforehand, meaning the methods cannot be used to find unknown reactions. TSDiff only requires the connectivity of the molecular graph of the reactant, while OA-ReactDiff requires the exact 3D geometry.
Second, both methods require data of transition states in combination with reactants. Acquiring data on transition states is expensive, and the largest existing datasets only contain on the order of $10^5$ samples. This means that in practice, these methods are severly data constrained.

### 3.2 Doob's Lagrangian: A Sample-Efficient Variational Approach to Transition Path Sampling

**Method and impact** Doob's Langrangion instead proposes a data-free variational approach for transition path sampling (TPS). Given two meta-stable states of molecular systems, the method can be used to sample the full posterior distribution over transition paths instead of just the transition state. For Brownian motion diffusion processes, like underdamped or overdamped langevin in molecular dynamics, conditioning on the endpoints can be achieved by Doob's h-transform. They approach solving for Doob's h-transform via a least action principle where, by defining a Lagrangian action that can be minimized. To be tractable the paths are parametrized by a mixture of Gaussians.

**Limitations** Altough Doob's Langrangian removes the need for data, it trades it for an expensive optimization that has to be performed for each transition (pair of reactant and product). They report that transition path sampling for alanine dipeptide requires 38-50 million energy function evaluations. Even though Doob's Langrangian samples the full transition path ensemble and NEB Henkelman et al. (2000) only the minimum energy path, NEB requires only on the order of 100 to 1000 energy function evaluations. This high cost makes Doob's Langrangian infeasable in practice.
There is currently no way that Doob's Langrangian can transfer across systems, i.e. it has to be retrained for each reactant-product-pair in each molecule. In contrast to a method like DEM, Doob's Langrangian cannot make use of preexisting data to speedup the optimization.

## 4 Research Vision

Given these the progress in the field, we end with a personal view that attempts to unify past and future developments.

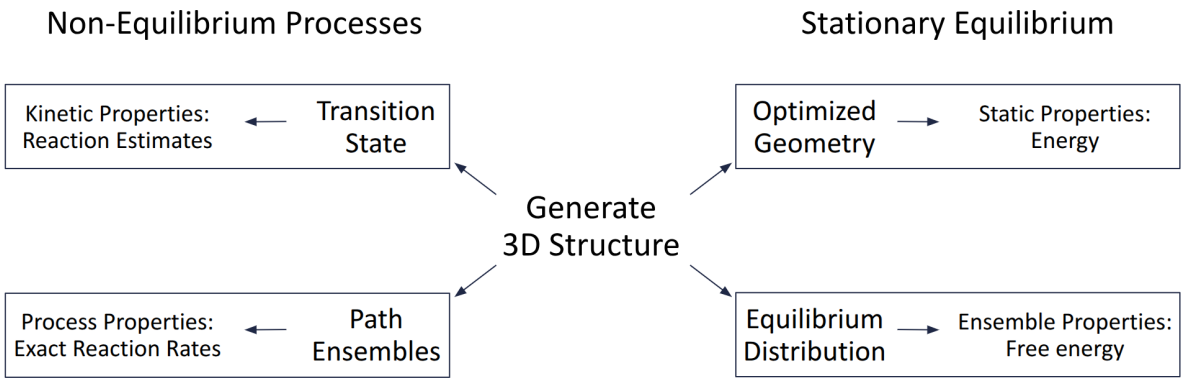


Figure 1: Personal, unifying view on generative modeling for molecular simulation. The field focuses from static geometries, to equilibrium ensembles, to transition states, and finally non-equilibrium path ensembles

Early work in 3D generative modeling focused on generating single structures. This includes the famous example of AlphaFold, which essentially predicts the lowest energy state (fold) of a protein. captures much or can be used as initial guess for optimization or simulation.
More recently the community focused on sampling from the equilibrium distribution, generating a range of structures according to their probability. Useful for computing expectation values The works reviewed in the first part of this paper Sec. 2 fall into this second category.
The third category of transitions states methods predicts a single structure that can approximately characterize a process. Works in Sec. 3 as well as our proposal below belong to this group. Focusing on the transition state is justified, as it gives mechanistic insights, is tractable, and often valid approximation that can be used as an initial guess to optimize a full path ensemble. Similar to predicting low-energy structures (first category), predicting transition states follows the idea that much of a system's information is summarized in the critical points (minima and saddle points) of the potential energy surface.
The equilibrium distribution is a convenient description, since all thermodynamics properties become time-independent. Many systems can, at least approximately, be described as being in equilibrium.

Nevertheless transition state theory (TST) and minimum energy paths (MEP) are equilibrium concepts, while transition path sampling (TPS) handles full non-equilibrium dynamics.
In this review arguably only Doob's Langrangian addresses the fourth category of full non-equilibrium dynamics by sampling path ensembles.
This leaves the question: which direction appears as the most fruitful for research? We argue for focusing on the simulation of (non-equilibrium) processes.
Finding and generating transition states remains an open problem which would be highly useful to the chemistry community if solved. In addition there a likely many techniques from generating individual or ensembles of low-energy structures that transfer to transition structures.
In the long run the most general description is given by path ensembles. If other parts of the sampling pipeline, like machine learning force fields, continue to become cheaper and more accurate, this level of accuracy might be desirable. There is ongoing work in the field to efficiently learn to generate trajectories Bartosh et al. (2025), and improve representations of paths Ramakrishnan et al. (2025), that might make sampling non-equilibrium dynamics with path ensembles feasible.

### 4.1 Proposal: Neural Transition State Sampling

We end by presenting the proposal for an ongoing work that addresses the problem of finding transition states.
The proposed method targets applications in both materials science (catalyst design) and drug discovery (small organic molecules). Catalyst design in particular requires identifying a diverse set of transitions. Beyond achieving the target chemical reaction, researchers must address side reactions that produce byproducts, leading to catalyst poisoning where accumulated impurities render the catalyst ineffective. To then compute transition rates, the minima can be derived from the transition state using intrinsic reaction coordinate (IRC) methods.
The proposed method addresses several key objectives not achieved by previous work. The approach should (1) transfer across molecular systems (2) identify a diverse set of unknown transitions (3) without requiring prior knowledge of reaction coordinates or stable minima (4) utilize available experimental or computational data when present (5) while maintaining the ability to learn on new systems without existing or poor data.
Our approach builds on neural samplers that have shown to be able to generate diverse molecular structures Havens et al. (2025). Prior works target the Boltzmann distribution over physical energy, where the distribution is maximal where the energy is minimal. Instead we want to sample a different pseudo-energy that is minimal at the transition states. Mathematically, transition states are index-1 saddle points, meaning the curvature is positive (local maximum) along one direction, the minimum energy path, and negative in all others. A pseudo-potential minimal at transition states $U$ could trivially be achieved by incorporating second-order information about the physical energy $E$

$$U(x) = \begin{cases} -\lambda_1\lambda_2 & \text{if } \lambda_1\lambda_2 > 0 \\ |\nabla E| & \text{otherwise} \end{cases} \tag{13}$$

where $\lambda_1$ and $\lambda_2$ are the two smallest eigenvalues of the Hessian $Hv_i(x) = \lambda_i v_i(x)$, $\lambda_1 < \lambda_2, \ldots, \lambda_N$. expensive but tractable with machine learning force fields using automatic differentiation Yuan et al. (2024). Since methods like Adjoint Sampling Havens et al. (2025) only require the force but not the potential, one can also learn a vector field for which no closed-form global potential exists E & Zhou (2010); Gu & Zhou (2018)

$$\dot{x} = -\nabla E + 2\langle \nabla E, v_1(x)\rangle v_1(x), \tag{14}$$

where $v_1(x)$ is the eigenvector associated with the smallest eigenvalue of the Hessian $Hv_1(x) = \lambda_1 v_1(x)$ and $\langle \cdot, \cdot \rangle$ is the inner product.
Note that we are not interested in sampling from the distribution itself, but use the Boltzmann distribution as an exploration mechanism. To only sample close to the modes, the transition states, we can anneal down the temperature. To preserve exploration on new systems we can condition on the temperature during training.
System transferability is achieved by conditioning the sampler on the atomic composition and connectivity of each molecular system. For large molecules containing numerous potential transitions, the approach can optionally limit scope by considering only transitions within specified energy ranges

or focusing on reactions around known connectivity patterns using biasing forces on chemical bonds Havens et al. (2025). The training process incorporates both supervised learning on available transition state data Zhao et al. (2023) using bridge matching Shi et al. (2023) and unsupervised exploration in regions where data remains scarce Havens et al. (2025).
Existing benchmarks probe individual transition states, but fail to cover the important task of discovering all relevant transitions within a system. This highlights the need to develop an adequate benchmark, with ground truth through systematic enumeration methods Maeda et al. (2016). Evaluation metrics will assess both coverage and quality of transition state predictions. Coverage metrics measure the diversity of discovered transitions relative to the complete set of possible pathways, while quality metrics evaluate the accuracy of predicted transition state geometries and energies compared to reference calculations.
This framework ensures that the method achieves both comprehensive exploration and reliable prediction accuracy necessary for practical applications.

# 5 Conclusion

In this paper, we surveyed recent advances in generative models that sample the equilibrium (Boltzmann) distribution as well as rare transition events.
Traditional simulation approaches through molecular dynamics or Monte Carlo methods face fundamental limitations due to long autocorrelation times and inadequate exploration of high-barrier regions. These constraints have driven the development of generative models that can efficiently generate independent samples in parallel while exploring complex potential energy surfaces.
While equilibrium sampling presents greater computational tractability, it provides limited insights into the dynamic processes that govern molecular function. Chemical and biological processes are fundamentally characterized by transition pathways and transition states rather than equilibrium configurations. These transition states represent rare events under equilibrium distributions, which means that traditional sampling approaches observe transitions infrequently, resulting in biased statistical estimates of process rates.
Existing non-machine learning methods for rare events often incur a high cost for each transition of interest or require extensive knowledge of the system to guide the optimization. Machine learning methods can often be expensive to train, but have the hope of transferring across systems. While machine learning approaches often require expensive initial training phases, they offer the potential for transferability across multiple molecular systems, providing better long-term computational efficiency. However, current machine learning methods for transition sampling have not yet achieved practical utility due to their reliance on either prohibitively expensive optimization procedures or requirements for difficult-to-obtain transition pathway data.
To address these gaps, we proposed a neural sampling framework for learning to sample a wide range of transition states. The method can incorporate data in form of (estimates of) transition states or use second-order (Hessian) information from machine learning force fields Yuan et al. (2024) to guide the sampling.
Promising avenues for future work could focus on the exploration chemical space by using the latent space of learned samplers Noé et al. (2019); Plainer et al. (2023). A possible approach could enhance replica exchange sampling with learnable transport maps between chains Zhang et al. (2025).
Ultimately, the most complete description of non-equilibrium processes is given by path ensembles. Thus, future objectives for the field should include the development of methods capable of sampling distributions over trajectories in a tractable manner.

# Acknowledgement

A.B. thanks Alán Aspuru-Guzik, Nandita Vijaykumar, Chris J. Maddison, and Varinia Bernales for helpful discussions. The proposed project is in collaboration with Nikolaj Rønne, Tejs Vegge, Arghya Bhowmik of the Technical University of Denmark (DTU) as well as Luca Thiede. Figure 1 was greatly inspired by a previous version from Austin H. Cheng.

# References


Tara Akhound-Sadegh, Jarrid Rector-Brooks, Joey Bose, Sarthak Mittal, Pablo Lemos, Cheng-Hao Liu, Marcin Sendera, Siamak Ravanbakhsh, Gauthier Gidel, Yoshua Bengio, Nikolay Malkin, and Alexander Tong. Iterated Denoising Energy Matching for Sampling from Boltzmann Densities. In *Forty-First International Conference on Machine Learning*, June 2024. 2.3

Abdulrahman Aldossary, Jorge Arturo Campos-Gonzalez-Angulo, Sergio Pablo-García, Shi Xuan Leong, Ella Miray Rajaonson, Luca Thiede, Gary Tom, Andrew Wang, Davide Avagliano, and Alán Aspuru-Guzik. In Silico Chemical Experiments in the Age of AI: From Quantum Chemistry to Machine Learning and Back. *Advanced Materials*, 36(30):2402369, 2024. ISSN 1521-4095. doi: 10.1002/adma.202402369. 2, 2.1

Grigory Bartosh, Dmitry Vetrov, and Christian A. Naesseth. SDE Matching: Scalable and Simulation-Free Training of Latent Stochastic Differential Equations, February 2025. 4

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. 2.2

Gavin E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721–2726, September 1999. doi: 10.1103/PhysRevE.60.2721. 3

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint Matching: Fine-tuning Flow and Diffusion Generative Models with Memoryless Stochastic Optimal Control. In *The Thirteenth International Conference on Learning Representations*, October 2024. 2.4

Chenru Duan, Yuanqi Du, Haojun Jia, and Heather J. Kulik. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature Computational Science*, 3 (12):1045–1055, December 2023. ISSN 2662-8457. doi: 10.1038/s43588-023-00563-7. 3.1

Weinan E and Eric Vanden-Eijnden. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annual Review of Physical Chemistry*, 61(Volume 61, 2010):391–420, May 2010. ISSN 0066-426X, 1545-1593. doi: 10.1146/annurev.physchem.040808.090412. 3

Weinan E and Xiang Zhou. The Gentlest Ascent Dynamics. *Chaos*, 2010. doi: 10.48550/ARXIV.1011.0042. 4.1

Henry Eyring. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*, 3(2):107–115, February 1935. ISSN 0021-9606. doi: 10.1063/1.1749604. 3

Daan Frenkel and Berend Smit. *Understanding Molecular Simulation*. Elsevier, 2002. ISBN 978-0-12-267351-1. doi: 10.1016/B978-0-12-267351-1.X5000-7. 2

Shuting Gu and Xiang Zhou. Simplified gentlest ascent dynamics for saddle points in non-gradient systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(12):123106, December 2018. ISSN 1054-1500. doi: 10.1063/1.5046819. 4.1

Aaron J. Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Daniel S. Levine, Brandon M. Wood, Bin Hu, Brandon Amos, Brian Karrer, Xiang Fu, Guan-Horng Liu, and Ricky T. Q. Chen. Adjoint Sampling: Highly Scalable Diffusion Samplers via Adjoint Matching. In *Frontiers in Probabilistic Inference: Learning Meets Sampling*, April 2025. 2.4, 4.1, 4.1, 4.1

Jiajun He, Yuanqi Du, Francisco Vargas, Dinghuai Zhang, Shreyas Padhy, RuiKang OuYang, Carla Gomes, and José Miguel Hernández-Lobato. No Trick, No Treat: Pursuits and Challenges Towards Simulation-free Training of Neural Samplers, April 2025. 2.3

Jérôme Hénin, Tony Lelièvre, Michael R. Shirts, Omar Valsson, and Lucie Delemotte. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living Journal of Computational Molecular Science*, 4(1):1583–1583, December 2022. ISSN 2575-6524. doi: 10.33011/livecoms.4.1.1583. 3

Graeme Henkelman, Blas P. Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*, 113 (22):9901–9904, December 2000. ISSN 0021-9606. doi: 10.1063/1.1329672. 3, 3.1, 3.2

C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*, 78(14): 2690–2693, April 1997. doi: 10.1103/PhysRevLett.78.2690. 3

Christopher Jarzynski. Rare events and the convergence of exponentially averaged work values. *Physical Review E*, 73(4):046105, April 2006. doi: 10.1103/PhysRevE.73.046105. 1, 3

Seonghwan Kim, Jeheon Woo, and Woo Youn Kim. Diffusion-based generative AI for exploring transition states from 2D molecular graphs. *Nature Communications*, 15(1):341, January 2024. ISSN 2041-1723. doi: 10.1038/s41467-023-44629-6. 3.1

Leon Klein and Frank Noe. Transferable Boltzmann Generators. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, November 2024. 2.2

Leon Klein, Andreas Krämer, and Frank Noe. Equivariant flow matching. *Advances in Neural Information Processing Systems*, 36:59886–59910, December 2023. 2.2

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, October 2002. doi: 10.1073/pnas.202427399. 3

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*, September 2022. 2.2

Satoshi Maeda, Yu Harabuchi, Makito Takagi, Tetsuya Taketsugu, and Keiji Morokuma. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *The Chemical Record*, 16(5):2232–2248, 2016. ISSN 1528-0691. doi: 10.1002/tcr.201600043. 3, 4.1

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, September 2019. doi: 10.1126/science.aaw1147. 2.1, 5

Baron Peters, Andreas Heyden, Alexis T. Bell, and Arup Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of Chemical Physics*, 120(17):7877–7886, May 2004. ISSN 0021-9606. doi: 10.1063/1.1691018. 3

Michael Plainer, Hannes Stark, Charlotte Bunne, and Stephan Günnemann. Transition Path Sampling with Boltzmann Generator-based MCMC Moves. In *NeurIPS 2023 AI for Science Workshop*, October 2023. 5

Lawrence R. Pratt. A statistical method for identifying transition states in high dimensional problems. *The Journal of Chemical Physics*, 85(9):5045–5048, November 1986. ISSN 0021-9606. doi: 10.1063/1.451695. 3

Kalyan Ramakrishnan, Lars L. Schaaf, Chen Lin, Guangrun Wang, and Philip Torr. Implicit Neural Representations for Chemical Reaction Paths, February 2025. 4

Lorenz Richter and Julius Berner. Improved sampling via learned diffusions, May 2024. 2.1

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 9323–9332. PMLR, July 2021. 2.2, 2.3

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger Bridge Matching. *Advances in Neural Information Processing Systems*, 36:62183–62223, December 2023. 2.4, 4.1

Marta Skreta, Lazar Atanackovic, Joey Bose, Alexander Tong, and Kirill Neklyudov. The Superposition of Diffusion Models Using the Itô Density Estimator. In *The Thirteenth International Conference on Learning Representations*, October 2024. 2.2

Charlie B. Tan, Joey Bose, Chen Lin, Leon Klein, Michael M. Bronstein, and Alexander Tong. Scalable Equilibrium Sampling with Sequential Boltzmann Generators. In *Frontiers in Probabilistic Inference: Learning Meets Sampling*, April 2025. 2.2

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, May 2018. 2.2

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, November 2023. ISSN 2835-8856. 2.2

Steyn van Leeuwen, Alberto Pérez de Alba Ortíz, and Marjolein Dijkstra. A Boltzmann generator for the isobaric-isothermal ensemble, May 2023. 2.5

Francisco Vargas, Will Sussman Grathwohl, and Arnaud Doucet. Denoising Diffusion Samplers. In *The Eleventh International Conference on Learning Representations*, September 2022. 2.1

Peter Wirnsberger, Borja Ibarz, and George Papamakarios. Estimating Gibbs free energies via isobaric-isothermal flows, September 2023. 2.5

Eric C.-Y. Yuan, Anup Kumar, Xingyi Guan, Eric D. Hermes, Andrew S. Rosen, Judit Zádor, Teresa Head-Gordon, and Samuel M. Blau. Analytical ab initio hessian from a deep learning potential for transition state optimization. *Nature Communications*, 15(1):8865, October 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-52481-5. 4.1, 5

Leo Zhang, Peter Potaptchik, Jiajun He, Yuanqi Du, Arnaud Doucet, Francisco Vargas, Hai-Dang Dau, and Saifuddin Syed. Accelerated Parallel Tempering via Neural Transports, May 2025. 5

Qiyuan Zhao, Sai Mahit Vaddadi, Michael Woulfe, Lawal A. Ogunfowora, Sanjay S. Garimella, Olexandr Isayev, and Brett M. Savoie. Comprehensive exploration of graphically defined reaction spaces. *Scientific Data*, 10(1):145, March 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-02043-z. 4.1

Paul M. Zimmerman. Single-ended transition state finding with the growing string method. *Journal of Computational Chemistry*, 36(9):601–611, 2015. ISSN 1096-987X. doi: 10.1002/jcc.23833. 3

# 6 Guidelines

Source: PhD Handbook 2023/2024

## 6.1 Guideline Summary

- select 5–10 research papers (...) in one research area
- This research area need not correspond to the student's eventual choice of PhD topic
- the student will be examined on the points (a) through (e)
  - (a) apply basic tools, know what they don't know (b) filter significant research contributions (c) relate papers to another (d) critique research methods (e) identify limitations in results
- it is not expected that the student master the majority of the relevant literature
- the student will prepare a short position paper (...) on points (c)–(e)
  - (c) relate papers to another (d) critique research methods (e) identify limitations in results

## 6.2 Relevant Excerpts

**5. Checkpoints and Supervisory Committee Meetings** Committee recommendations: After each supervisory committee meeting, the chair of the supervisory committee will provide written feedback to the student (through the Graduate Office) and the student will be invited by the Graduate Office to respond to this feedback. In addition, one of the following results will be provided:

1. Pass: A pass may be accompanied by constructive feedback and/or suggestions for activity in the next session(s).
2. Conditional Pass: The student is given one or more concrete tasks to complete by a specific deadline (no more than 12 months later). The tasks and the deadline are also communicated to the Graduate Office. The meeting chair is responsible for reporting to the Graduate Office whether or not the student has cleared the conditions by the deadline. If the student fails to clear the conditions by the deadline, their progress will be considered unsatisfactory.
3. Fail (with the option to repeat): The student is not considered to be making satisfactory academic progress and must hold another supervisory committee meeting within 6 months.
4. Fail (no option to repeat): The committee recommends that the student must either withdraw from the program or have their registration terminated. This result is possible only for students who were not considered to be making satisfactory academic progress prior to the meeting. The Associate Chair, Graduate Studies will review such a recommendation.

**5.2 Qualifying Oral Examination** Students, working with their supervisor, should select 5–10 research papers to be emphasized at the Qualifying Oral. These should be important papers in one research area of relevance to CS. This research area need not correspond to the student's eventual choice of PhD topic — students need not be committed to a thesis topic at this stage. In relation to the selected papers, the student will be examined on the points (a) through (e) listed in Section 1 (6.2). It is expected that students will have read and understood more than just the selected papers, but it is not expected that the student master the majority of the relevant literature at the time of this exam. In order to help focus the initial questioning, the student will prepare a short position paper (about 4,000 words or 8 single spaced pages in a reasonable font) on points (c)–(e), as outlined in section 1 (6.2) above. If the student has begun to investigate this area themselves, then they are welcome to briefly describe their progress so far. In addition, it is the student's option to discuss the expected overall scope of the questioning with their supervisory committee prior to the exam. This paper should be submitted to their committee in at least one week in advance of the meeting.

**Section 1** Specific foundational skills to be developed by a PhD candidate include these:

- (a) The ability to apply the basic tools of the field in potentially new ways, along with understanding what they know and what they have yet to learn.
- (b) The ability to select significant research contributions from a larger set of published papers, and justify that selection (for example, on the basis of the significance of the results or the novelty of the approach).
- (c) The ability to relate the papers to one another, and to other research in the literature.
- (d) The ability to critique the research methods used in these papers, including the strengths and weaknesses of these methods and likely threats to validity, whether these are acknowledged in the

papers or not.

- (e) The ability to identify limitations of the results (and possibly errors) reported in the papers, along with their implications.