Pink Noise Is All You Need: Colored Noise Exploration in Deep RL

Onno Eberhard[1] Jakob Hollenstein[2,1] Cristina Pinneri[3,1] Georg Martius[1]

[1]Max Planck Institute for Intelligent Systems [2]Universität Innsbruck [3]Max Planck ETH Center for Learning Systems

Abstract

- Setting: Off-Policy reinforcement learning for **continuous control**
- Exploration is commonly performed by adding random perturbations to the actions or, equivalently, by sampling actions from a stochastic policy.
- This **white noise** exploration is often not sufficient to find high reward regions
- Strongly temporally correlated alternatives like Ornstein-Uhlenbeck (OU) noise, which try to tackle this issue, can inhibit learning when not necessary
- We examine the effectiveness of **colored noise** of intermediate temporal correlation
- Our results show that **pink noise** significantly outperforms white noise and OU noise across many tasks, and should be preferred as the **default choice** for action noise

Action Noise for Exploration

In off-policy RL, action noise ($\varepsilon_t \sim \mathcal{N}(0, I)$) may be added to a deterministic policy:

$$a_t = \mu(s_t) + \sigma\varepsilon_t,$$

or used for sampling from a stochastic policy $\pi(a \mid s) = \mathcal{N}(a \mid \mu(s), \mathrm{diag}(\sigma^2(s)))$:

$$a_t = \mu(s_t) + \sigma(s_t) \odot \varepsilon_t.$$

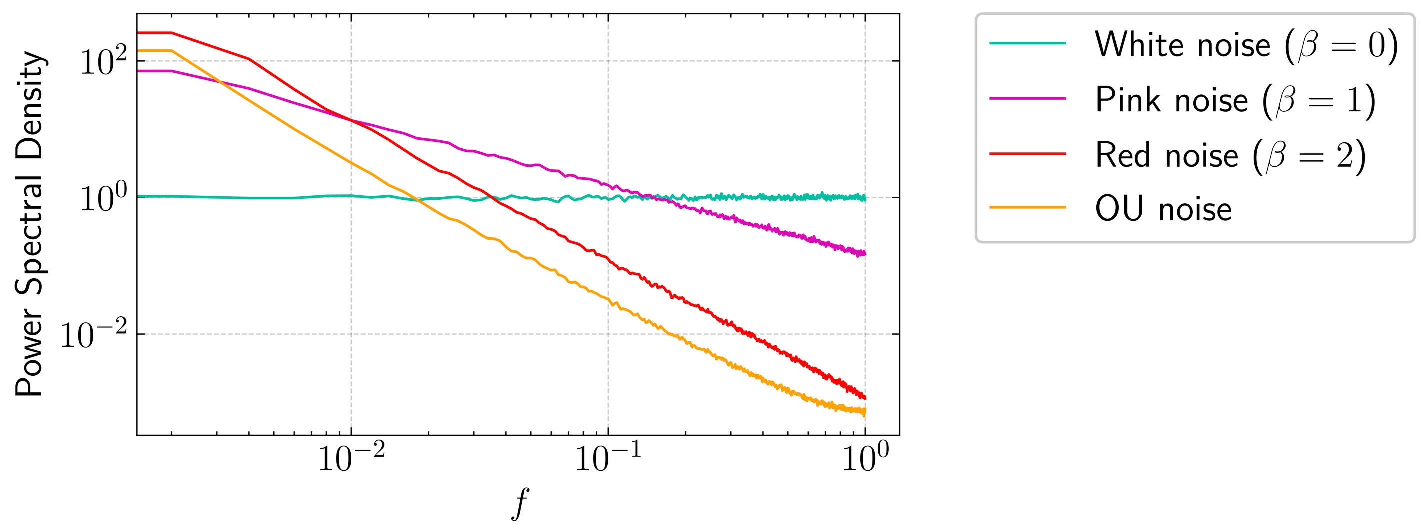
In both these cases, the noise signal $\varepsilon_{1:T}$ has no temporal correlation and is called **white noise**. Some tasks require stronger exploration and are better served by temporally correlated noises like Ornstein-Uhlenbeck (OU) noise:

$$\varepsilon_{t+1} \sim (1 - \theta\Delta t)\varepsilon_t + \sigma\mathcal{N}(0, \Delta t).$$

For many tasks OU noise is too strongly correlated → idea: **intermediate correlation**

Colored Noise

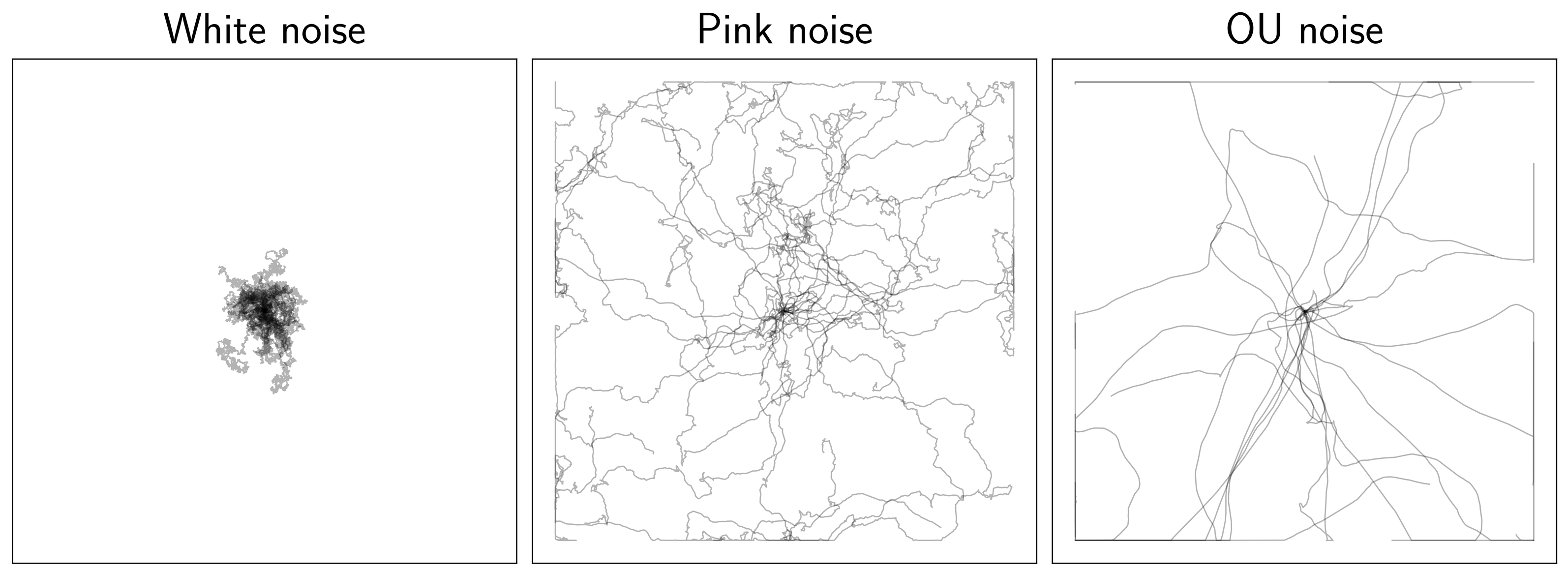
A stochastic process is called **colored noise** with color parameter β, if signals $\varepsilon(t)$ drawn from it have the property that $|\hat{\varepsilon}(f)|^2 \propto f^{-\beta}$, where $\hat{\varepsilon}(f)$ denotes the Fourier transform of $\varepsilon(t)$ and $|\hat{\varepsilon}(f)|^2$ is called the power spectral density.



Colored noise can be cheaply generated, and can

- interpolate between uncorrelated (white) and strongly correlated (red) noise,
- has already been shown to be effective in model-based reinforcement learning [1].

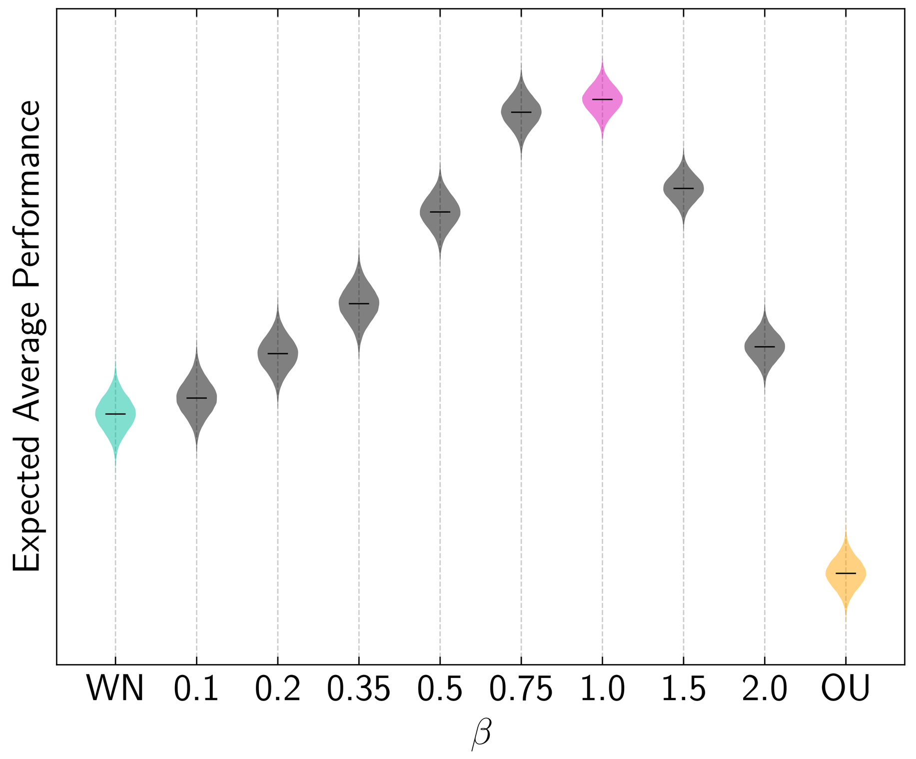
Intuition

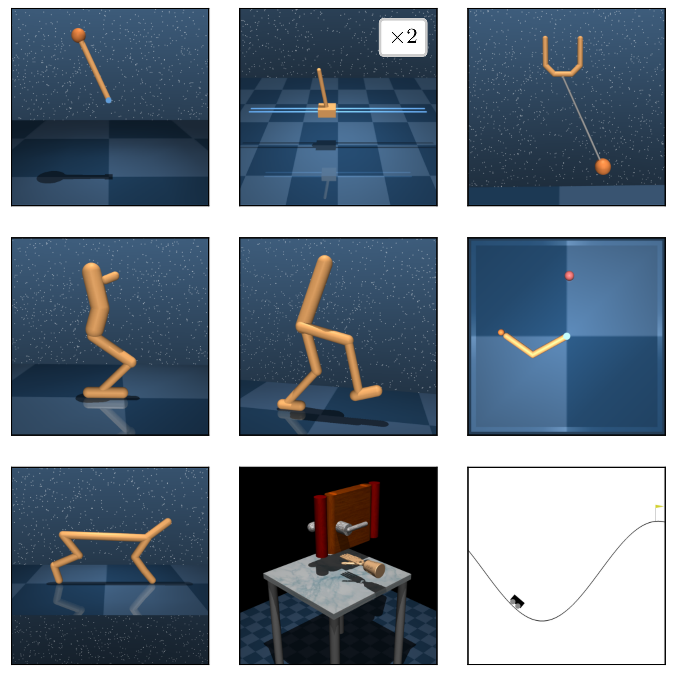


Trajectories on the bounded integrator environment ($c = 250$).

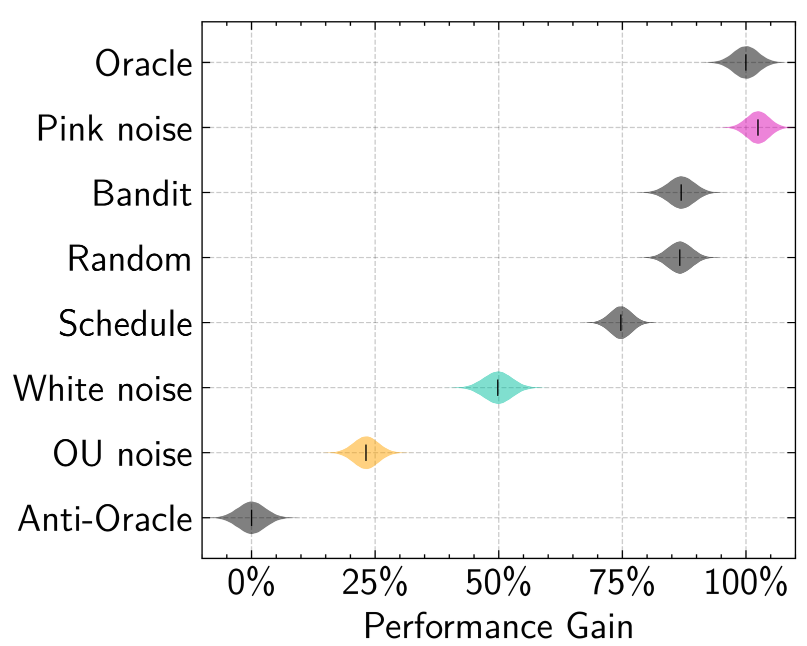
Experiments

We test 10 noise types on 10 benchmark environments using MPO [2] and SAC [3].





Pink noise ($\beta = 1$) significantly outperforms white noise (WN) and Ornstein-Uhlenbeck (OU) noise when performance is averaged across all benchmark environments. On 8/10 tasks, there is no significant difference between pink noise and the best noise type.



Can we be find a better strategy?

- Scheduling from $\beta = 2$ to $\beta = 0$
- Adapting β online to the task using a bandit algorithm

Results:

- Pink noise outperforms all alternatives significantly
- Pink noise performs on par with an oracle (empirically choosing best β for each task)

Pink noise is a **better default** than white noise and Ornstein-Uhlenbeck noise.

The Power of Pink

What makes pink noise a better default than white noise or OU noise? We examine this question using two simple environments which mirror common dynamics:

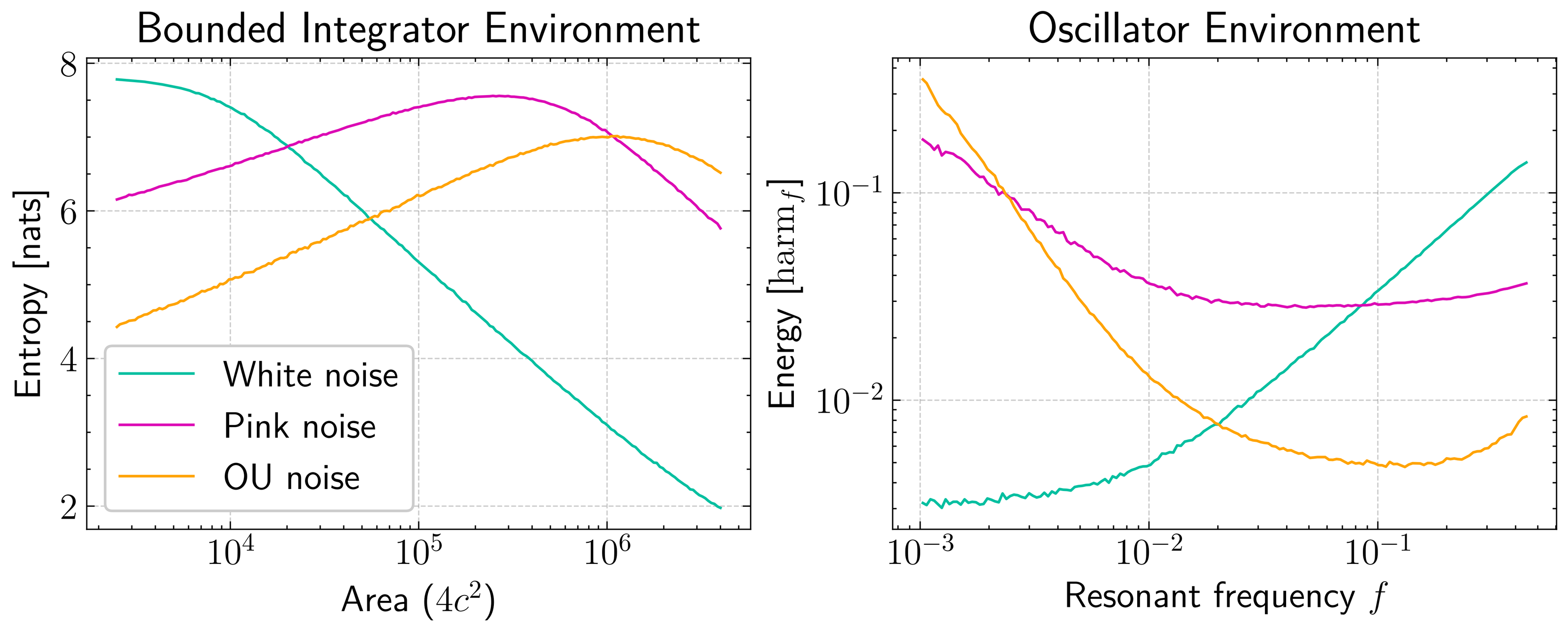
1. A bounded integrator:

$$\boldsymbol{s}_{t+1} = \mathrm{clip}(\boldsymbol{s}_t + \boldsymbol{a}_t, -c\mathbf{1}, +c\mathbf{1}) \qquad \rightarrow \text{Parameterized by area } (4c^2)$$

2. A harmonic oscillator:

$$\ddot{x} = \frac{F}{m} - \frac{k}{m}x \qquad \rightarrow \text{Parameterized by resonant frequency } f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

We now vary the parameters (c, f) over the complete sensible range (for episode lengths of $T = 1000$ and noise with $\mathrm{var}[\varepsilon_t] = 1$) and measure the quality of exploration.

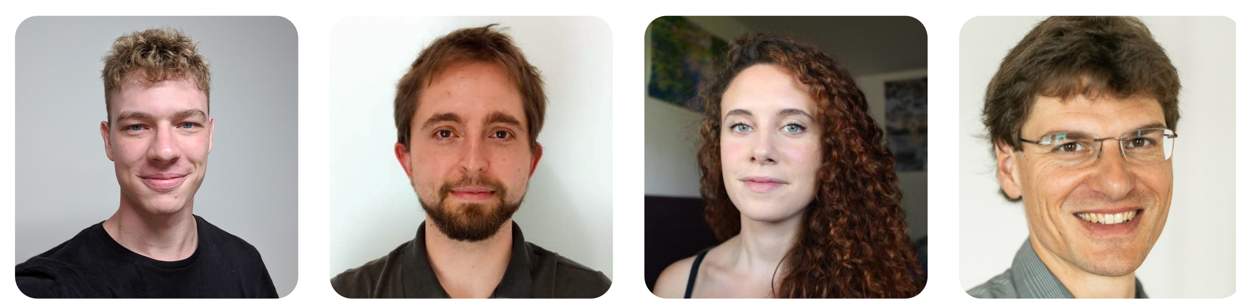


Takeaway: The intermediate temporal correlation makes pink noise **more general**.

- It is less sensitive to the environment parameterization than white noise / OU noise. → If the parameterization (e.g. c or f) is unknown, pink noise is the best choice.
- This explains the good average performance on the benchmark experiments.

Conclusion

We recommend **pink noise** as the **default choice** for action noise in reinforcement learning for continuous control. `pip install pink-noise-rl`



References

[1] Cristina Pinneri et al. *Sample-efficient Cross-Entropy Method for Real-time Planning*. CoRL 2020.
[2] Abbas Abdolmaleki et al. *Maximum a Posteriori Policy Optimisation*. ICLR 2018.
[3] Tuomas Haarnoja et al. *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor*. ICML 2018.