

# Parallel Denoising via Differential Noise Estimation in Diffusion Models

---

## **Abstract — "*Parallel Denoising via Differential Noise Estimation in Diffusion Models*"**

---

Denoising Diffusion Probabilistic Models (DDPM) have emerged as a powerful framework for generative modeling, but suffer from inherently sequential sampling procedures due to their autoregressive reverse process. In this work, we propose a novel formulation of the reverse process based on **differential noise estimation**, wherein the generative trajectory from pure noise to data is reconstructed by learning and subtracting a series of **incremental noise corrections**.

We demonstrate that, under the standard DDPM framework, the forward marginal distributions admit an exact closed-form expression for the difference between successive steps  $x_t - x_{t-1}$ , which can be interpreted as a structured residual noise term dependent on the data and latent variables. Leveraging this, we train a neural network  $\delta_\theta(x_T, t)$  to directly predict these differential noise terms **using only the final noise vector**  $x_T$  and a timestep embedding  $t$ , without conditioning on the intermediate states  $x_t$ .

This formulation leads to a **reverse sampling process that is both mathematically consistent and parallelizable**, as all residual noise corrections can be predicted independently and accumulated to obtain the final sample  $x_0$ . We provide a rigorous probabilistic analysis showing that this approach corresponds to a discretization of a deterministic flow over the latent space, and is solvable under Gaussian priors.

Empirical results on synthetic and real datasets demonstrate competitive generation quality and significantly improved sampling speed, confirming that the proposed method retains the expressivity of classical DDPMs while eliminating the need for sequential inference. This opens the path toward **truly parallel diffusion-based generative models**.

## 1. Introduction

---

Denoising Diffusion Probabilistic Models (DDPMs) have emerged as a powerful class of generative models, capable of synthesizing high-quality samples by learning to reverse a forward noising process. However, a major bottleneck in these models is their reliance on a sequential denoising trajectory, which limits inference speed and parallelization.

We propose a novel reformulation of the DDPM reverse process that replaces the step-wise conditional estimation of sample states with the estimation of a sequence of stochastic noise differentials, all predicted directly from the final noise vector  $x_T$  and diffusion timestep  $t$ . This reformulation allows us to view the reverse process as a discrete Langevin-like stochastic flow, enabling efficient and potentially parallel sampling while maintaining generative diversity.

## 2. Background

---

In the standard DDPM framework, the forward process is defined as a Markov chain:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I})$$

with  $\beta_t \in (0, 1)$  denoting a variance schedule. The marginal distribution of  $x_t$  conditioned on the original data  $x_0$  is given in closed form as:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I})$$

where  $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$ .

The reverse process is parameterized as:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

and the model is trained to minimize a variational bound or an equivalent denoising loss:

$$\mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2]$$

### 3. Differential Noise Estimation as a Reverse Flow

---

We redefine the reverse process by modeling the stochastic difference between successive noisy states:

$$\delta_t := x_t - x_{t-1}$$

This differential can be expressed analytically under the forward process as:

$$\delta_t = (\sqrt{\bar{\alpha}_t} - \sqrt{\bar{\alpha}_{t-1}})x_0 + (\sqrt{1 - \bar{\alpha}_t} - \sqrt{1 - \bar{\alpha}_{t-1}})\epsilon$$

which is a Gaussian random variable dependent on  $x_0$  and  $\epsilon$ .

We train a model  $\delta_\theta(x_T, t)$  to approximate the distribution of  $\delta_t$  by regressing its mean and variance conditioned only on the initial noise vector  $x_T$  and the timestep  $t$ . This yields:

$$\delta_\theta(x_T, t) = \mu_\theta(x_T, t) + \sigma_\theta(x_T, t) \cdot z, \quad z \sim \mathcal{N}(0, \mathbf{I})$$

This construction enables the reverse process to be interpreted as a discretized stochastic flow:

$$x_{t-1} = x_t - \delta_\theta(x_T, t)$$

### 4. Langevin Interpretation

---

By explicitly modeling  $\delta_\theta$  as a random variable with learnable mean and variance, we recover the structure of a Langevin reverse-time SDE:

$$dx = -\mu_\theta(x_T, t)dt + \sqrt{2\beta(t)}dW_t$$

where  $dW_t$  is a Wiener process and  $\beta(t)$  controls the diffusion scale. We approximate this flow using Euler-Maruyama discretization:

$$x_{t-1} = x_t - \mu_\theta(x_T, t) - \sigma_\theta(x_T, t) \cdot z_t$$

This formulation preserves the stochastic nature of generation, in contrast to deterministic ODE-based samplers, and allows direct control over the noise profile through  $\sigma_\theta$ .

Perfetto. Ecco la sezione formale per il tuo paper:

---

## On the Stochastic Nature of the Model

---

### Background: Langevin Dynamics in Diffusion Models

In traditional Denoising Diffusion Probabilistic Models (DDPM), the reverse generative process is theoretically grounded in **Langevin dynamics**, where one seeks to approximate samples from the data distribution via:

$$dx = \nabla_x \log p_t(x) dt + \sqrt{2} dW_t$$

where:

- $p_t(x)$  is the marginal distribution at time  $t$
- $dW_t$  is the standard Wiener process

In the discrete setting, this is approximated using a **finite-step stochastic update**:

$$x_{t-1} = x_t + \underbrace{\frac{1}{2} \nabla_x \log p_t(x_t) \cdot \Delta t}_{\text{score}} + \sqrt{\Delta t} \cdot z, \quad z \sim \mathcal{N}(0, I)$$

Thus, DDPMs leverage a learned score function  $\nabla_x \log p_t(x) \approx s_\theta(x, t)$  and perform **stochastic reverse diffusion**, driven by both gradient flow and injected Gaussian noise.

---

### Our Proposal: Differential Noise Modeling (DNM)

In our formulation, we propose to learn a mapping:

$$\delta_t = x_t - x_{t-1}$$

with the objective of reconstructing the data point  $x_0$  from a fully noised input  $x_T \sim \mathcal{N}(0, I)$  via:

$$x_0 = x_T - \sum_{t=1}^T \delta_\theta(x_T, t)$$

The key innovation lies in predicting **local differentials** rather than global denoising steps or pixel values.

---

### Stochastic Extension: Langevin-Inspired Delta Sampling

To reintroduce stochasticity akin to Langevin processes, each predicted step  $\delta_t$  can be interpreted as a **mean term** of a local Gaussian perturbation:

$$\delta_t \sim \mathcal{N}(\mu_\theta(x_T, t), \sigma_t^2 I)$$

and the cumulative generation becomes:

$$x_0 = x_T - \sum_{t=1}^T (\mu_\theta(x_T, t) + \sigma_t z_t), \quad z_t \sim \mathcal{N}(0, I)$$

This yields a **Langevin-like stochastic delta process**, which preserves:

- the **incremental nature** of the dynamics
- the **stochasticity of reverse steps**
- the **parallel sampling capacity** (since  $x_T$  is fixed and all  $t$  are independent)

By choosing the noise scale  $\sigma_t \propto \sqrt{\Delta t}$ , this structure approximates the discrete-time Langevin formulation. If  $\sigma_t = 0$ , we recover the **deterministic variant**.

## Implications

- The model transitions from a deterministic “ODE-like” residual model to a **stochastic SDE-based** formulation.
- This allows **multiple stochastic samples** from the same  $x_T$ , increasing expressiveness.
- The process is compatible with **likelihood estimation** via score-matching or ELBO analogs if extended further.

## Conclusion

Our model lies at the intersection between deterministic residual modeling and probabilistic Langevin-based diffusion. By explicitly modeling delta-noise steps and optionally injecting stochasticity, we preserve both **interpretability** and **generative diversity**, while enabling **efficient, parallel sampling**.

Vuoi che converta questa sezione in formato LaTeX o Markdown pronto da incollare nel paper?

## 5. Implications and Advantages

- **Parallelization**: Since all  $\delta_\theta(x_T, t)$  are conditioned only on  $x_T$ , sampling steps can be computed independently.
- **Stochasticity**: Incorporating Gaussian noise maintains sample diversity, unlike deterministic samplers.
- **Modularity**: The method cleanly separates the modeling of the noise from the integration path.
- **Flexibility**: By predicting both mean and variance, the method generalizes DDPM and DDIM as special cases.

## 6. Future Work

This framework opens several directions: efficient training via score distillation, integration with latent diffusion, and theoretical analysis of the consistency between the estimated  $\delta_\theta$  and the true reverse SDE trajectories.

## References

---

(References will be added later, including Ho et al. (2020), Song et al. (2020), and related Langevin SDE literature.)