Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bending Moment Synthesis by Integrating Cutting Mechanics with Diffusion Model for Tool Wear Prediction

Yu-Sheng Lin
Jian-Ping Jhuang
Meng-Shiun Tsai
mstsai0126@ntu.edu.tw

National Taiwan University https://orcid.org/0000-0003-2812-5975

Research Article

Keywords: Bending Moment, Data Augmentation, Diffusion Model, Tool Wear

Posted Date: July 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10341918/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Bending Moment Synthesis by Integrating Cutting Mechanics with Diffusion Model for Tool Wear Prediction

Yu-Sheng Lin [1], Jian-Ping Jhuang [2] and Meng-Shiun Tsai [3,*]

[1,2,3] Department of Mechanical Engineering, National Taiwan University, No. 1, Roosevelt Rd., Taipei 106216, Taiwan, R.O.C

***** Correspondence: mstsai0126@ntu.edu.tw

Abstract:

In smart manufacturing, accurate tool wear prediction is often limited by data deficiency and cross-condition adaptation. Conventional data-driven generative models, such as GANs, synthesize high-frequency dynamic signals without physical guidance. This may cause frequency shifts and negative transfer in downstream prediction tasks. To address this issue, this study proposes a cutting bending moment generation framework that integrates a mechanistic cutting bending moment model with a diffusion model. Different from conventional physics-informed or hybrid ML frameworks that usually impose physical equations through loss functions or simulation-assisted data generation, the proposed framework uses the ideal cutting bending moment as a physical input condition and injects the tool wear state through Feature-wise Linear Modulation (FiLM) during diffusion-based denoising. The proposed method uses a 1D-CNN U-Net as the denoising backbone. This design allows the model to generate cutting bending moment signals that preserve the main physical periodicity and wear-related signal variations. Cutting experiment results show that, using only 5% of real training data, the proposed method improves the coverage of wear-related features in the training set. In cross-condition tool wear prediction, the proposed method reduces the Root Mean Square Error (RMSE) by 60.45% compared with the baseline model trained only with real data. These results show that the proposed bending moment generation framework can support tool wear prediction under unseen cutting conditions and reduce the need for additional full-life tool wear experiments for tools with the same number of flutes.

Keywords: Bending Moment, Data Augmentation, Diffusion Model, Tool Wear

1. Introduction

In current industrial automation and smart manufacturing, accurate tool wear monitoring is paramount for ensuring machining quality and cost control. In precision machining processes, tool wear directly impacts the dimensional accuracy and surface quality of products [1]. Furthermore, it often leads to unexpected tool breakage, accounting for up to 20% of unplanned machine downtime [2].

In recent years, tool wear detection technologies have been extensively studied and can be broadly classified into direct and indirect methods based on data acquisition sources. Direct methods primarily employ machine vision to capture tool images, utilizing camera modules to directly quantify the location and severity of edge wear [3]. In contrast, indirect methods estimate tool wear by using machine or deep learning models to extract features from dynamic cutting signals, such as spindle current, cutting force, and vibration [4–6].

In current literature, indirect tool condition monitoring employs data-driven approaches. Manwar et al. [7] combined Long Short-Term Memory (LSTM) networks with cutting force signals to predict tool conditions. Chou et al. [8] proposed a Time-Frequency Markov Transition Field to transform time-domain data into two-dimensional images for tool state classification. Xu et al. [9] designed a novel multi-scale feature fusion architecture that integrates multi-source signals for tool wear prediction. However, the performance of these data-driven networks is constrained by the completeness and diversity of the training data. In a real-world machining environment, re-collecting comprehensive wear data for every new tool or varying cutting parameter incurs high costs. To address cross-condition generalization, transfer learning and domain adaptation strategies have been explored [10], although these approaches still require a sufficient amount of labeled target-domain data to achieve effective feature alignment. This data scarcity problem is widely recognized as a fundamental challenge in intelligent prediction and health management, where insufficient labeled samples significantly degrade model generalization under varying operating conditions [11]. To overcome these data acquisition challenges and improve cross-condition adaptability, recent studies have increasingly explored data augmentation and generative models to expand training datasets by synthesized signals.

Among various data augmentation techniques, Generative Adversarial Networks (GANs) are commonly applied to synthesize time-series signals. Shao et al. [12] proposed a GAN-based architecture that utilizes adversarial learning directly on sensor signals to generate virtual data for fault detection. Shah et al. [13] acquired cutting data by Acoustic Emission (AE) sensors and integrated GANs for data generation, subsequently employing a deep learning model to predict tool wear. Liu et al. [14] combined an LSTM model with a GAN to generate the AE sensor features, subsequently performing state prediction using a classifier. He et al. [15] constructed a GAN-based generative framework utilizing cutting vibration signals, incorporating an attention mechanism to develop a cross-domain generative network. Their downstream wear analysis verified the effectiveness of the synthesized data in improving the classification accuracy of various tool wear states. Additionally, many researchers transform 1D time-domain signals into 2D feature images, introducing architectures such as diffusion models [16] and GANs for data augmentation. The objective is to synthesize 2D images that preserve original time-domain characteristics by the restoration mechanisms of these image generative models. Liu et al. [17] transformed cutting signals into spectrograms and generated expanded image datasets using GANs, thereby enhancing the accuracy of downstream classification tasks. Chen et al. [18] converted cutting force signals into Symmetrized Dot Patterns (SDP) and applied a Denoising Diffusion

Probabilistic Model (DDPM) to learn and generate feature images across different wear states. Finally, an image classification network was utilized to predict the tool condition.

However, whether directly generating 1D waveforms using GANs or synthesizing 2D images with diffusion models, real dynamic signals covering a complete wear lifecycle remains difficult in TCM applications. Subject to severe data constraints, exclusively data-driven models fail to capture wear features. To address the dependency of purely data-driven models on real data, some researchers have introduced numerical simulation to assist in data generation. Zhu et al. [19] conducted simulated cutting experiments by Finite Element Method (FEM). They integrated these simulations with a small set of real cutting force data to generate augmented samples using a GAN architecture, which were then processed by a classifier for wear prediction. Although analytical cutting force models [20, 21] could directly derive cutting responses through mechanical mapping, they are unable to adapt to dynamic geometric variations induced by tool wear.

FEM numerical simulations can be used to generate data for different wear ranges. However, this approach usually requires considerable computational time and detailed boundary condition settings. Physics-Informed Neural Networks (PINNs) [22] have also been explored to combine physical laws with deep learning in structural health monitoring. These models usually add physical equations to the loss function as penalty terms. For high-frequency machining signals, this loss-based constraint may be difficult to optimize, and the generated signals may still deviate from the expected physical behavior. Existing literature employs cutting force as the monitoring indicator. However, the physical installation of force dynamometers is frequently and severely constrained by limited machine space and workpiece. In contrast, this study adopts the cutting bending moment as the monitoring feature, which can be analytically derived from cutting force components [23] and dynamically acquired using a smart tool holder with embedded sensors.

To address the limitations of theoretical modeling and the geometric variations caused by tool wear, this study proposes a cutting bending moment generation framework that combines a mechanistic cutting model with a diffusion model. The framework aims to support tool wear prediction for tools with the same number of flutes under different machining conditions. The framework uses the ideal cutting bending moment and the tool wear state as physical conditions. The diffusion model is built on a 1D-CNN U-Net backbone. A Feature-wise Linear Modulation (FiLM) block [24] is used to adjust the network features according to the input wear condition during the denoising process. This conditioning mechanism has been shown to be effective in 1D time-series modeling, where FiLM-based modulation allows CNN backbones to scale and shift feature channels based on external conditions [25]. Different from PINN-based methods, which usually add physical equations to the loss function as soft constraints, the proposed method uses the ideal bending moment as an input condition of the network. This setting helps the generated signals follow the main periodic structure of the cutting process. The synthesized data are then combined with real signals and used for downstream tool wear prediction.

2 Methodology

Table 1. Notation Table

Symbol	Definition	Unit / Remark
F_t	Tangential cutting force component acting on the cutting flute	N
F_n	Radial cutting force component acting on the cutting flute	N
F_a	Axial cutting force component acting on the cutting flute	N
K_{tc}, K_{nc}, K_{ac}	Tangential, radial, and axial cutting force coefficients	N/mm^2
K_{te}, K_{ne}, K_{ae}	Tangential, radial, and axial edge force coefficients	N/mm
$g(\phi)$	determine if the flute is within the cutting zone.	-
z	Axial depth	mm
b	discretized axial depth	mm
θ	Tool rotation angle	rad
$h(\phi)$	Instantaneous uncut chip thickness as a function of the flute engagement angle	mm
ϕ	Angular position	rad
ϕ_s	Entry angle of the cutting flute into the cutting zone	rad
ϕ_e	Exit angle of the cutting flute from the cutting zone	rad
N_t	Number of tool flutes	-
L_1	Moment arm associated with tangential and radial cutting force components	m
L_2	Moment arm associated with the axial cutting force component, corresponding to the tool radius in the bending moment model	m
B_x	Cutting bending moment in the x direction	Nm
B_y	Cutting bending moment in the y direction	Nm
M_{sim}	ideal bending moment generated from the mechanistic cutting model	Nm
w	Tool Wear state used as a conditional input	mm
t	Diffusion timestep used in the denoising process	-
γ, β	FiLM scaling function and shifting function	-
h_{out}	Output feature map after FiLM modulation	Tensor

As illustrated in Fig. 1, this study proposes a diffusion model based on the cutting bending moment. The architecture first employs a physics-guided strategy to address data scarcity in machining scenarios. The framework consists of two main components: a cutting bending moment model based on mechanistic cutting theory, and a conditional diffusion architecture with a Feature-wise Linear Modulation (FiLM) mechanism. On the right side of the architecture, real experimental data are segmented and combined with the synthesized cutting bending moment data for downstream tool wear prediction.

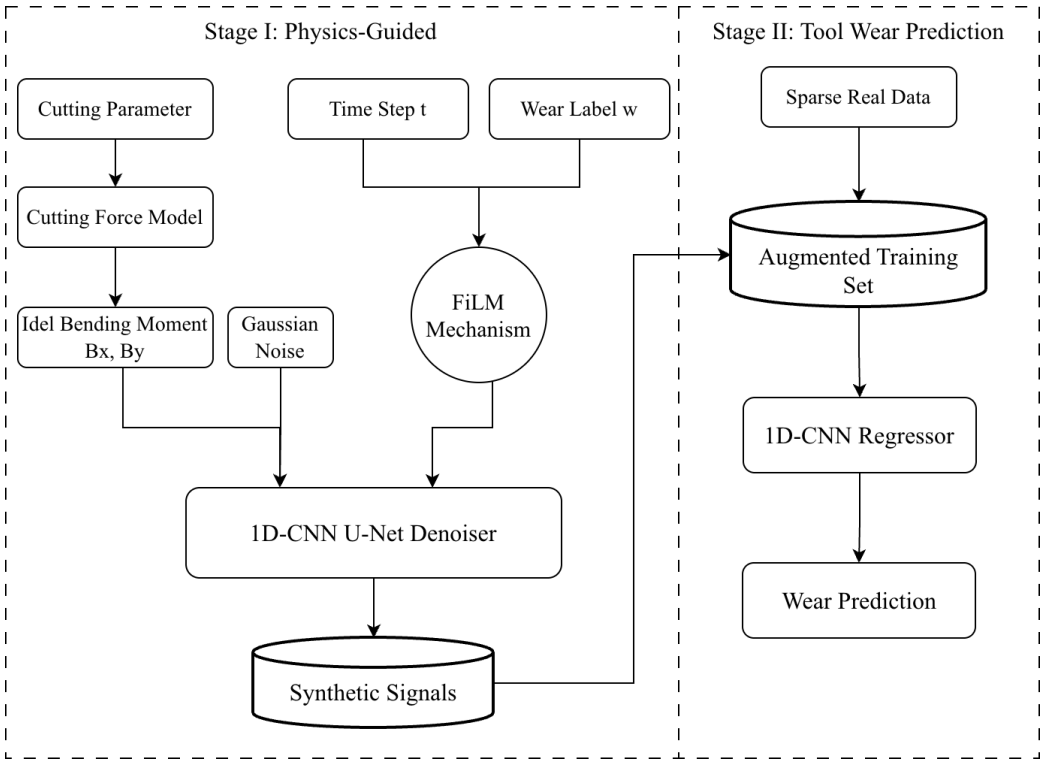


Fig. 1 Flowchart

2.1. Cutting Force and Bending Moment Model

Existing literature indicates that generative adversarial network (GANs) typically require large datasets to effectively capture and replicate the underlying data distribution. To ensure that the synthesized sensor signals follow the physical laws of machining processes rather than data-driven distributions, this study incorporates an analytical cutting force model [21] as a mechanistic prior. Unlike traditional generative approaches that start from random noise, this method uses the cutting bending moment based on cutting mechanics as a physical constraint to guide the diffusion process.

During the cutting process of a flat end mill, the instantaneous cutting forces acting on the j^{th} flute at axial depth z could be decomposed into tangential $F_{t,j}$, radial $F_{n,j}$, and axial $F_{a,j}$ components. As shown in (1), these forces depend on the cutting coefficients and the instantaneous chip thickness. Here, ϕ_j is the engagement angle of the j^{th} flute, and $h(\phi_j)$ is the instantaneous chip thickness. The cutting force coefficients (K_{tc}, K_{nc}, K_{ac}, K_{te}, K_{ne}, K_{ae}) are obtained from cutting experiments and are considered known constants.

$$\begin{aligned} dF_{t,j}(\phi_j, z) &= \left(K_{tc}h(\phi_j) + K_{te}\right)dz \\ dF_{n,j}(\phi_j, z) &= \left(K_{nc}h(\phi_j) + K_{ne}\right)dz \\ dF_{a,j}(\phi_j, z) &= \left(K_{ac}h(\phi_j) + K_{ae}\right)dz \end{aligned} \quad (1)$$

The bending moments (B_x, B_y) are calculated by multiplying the cutting forces by their corresponding moment arms. The total bending moment is obtained by integrating the product of these unit forces and their respective moment arms (L_1, L_2) along the axial depth z. Additionally, a switching function $g(\phi)$ is used to determine if the j^{th} flute is within the cutting zone. The derivation is shown in (2).

$$\mathrm{B_x}(\phi) = \sum_{\mathrm{j=1}}^{\mathrm{Nt}} \int_{\phi_s}^{\phi_e} \int_0^z \left[\left(\mathrm{F_{tj}}(\phi_j) + \mathrm{F_{nj}}(\phi_j)\right) \mathrm{L1} + \mathrm{F_{aj}}(\phi_j)\mathrm{L2}\right] \mathrm{g}(\phi_j) \cdot \mathrm{db} \cdot \mathrm{d}\phi_j$$

$$\mathrm{B_y}(\phi) = \sum_{\mathrm{j=1}}^{\mathrm{Nt}} \int_{\phi_s}^{\phi_e} \int_0^z \left[\left(\mathrm{F_{tj}}(\phi_j) - \mathrm{F_{nj}}(\phi_j)\right) \mathrm{L1} + \mathrm{F_{aj}}(\phi_j)\mathrm{L2}\right] \mathrm{g}(\phi_j) \cdot \mathrm{db} \cdot \mathrm{d}\phi_j \quad (2)$$

$$g(\phi_j) = \begin{Bmatrix} 1 & \phi_s \le \phi_j \le \phi_e \\ 0 & \phi_j < \phi_s, \phi_j > \phi_e \end{Bmatrix}$$

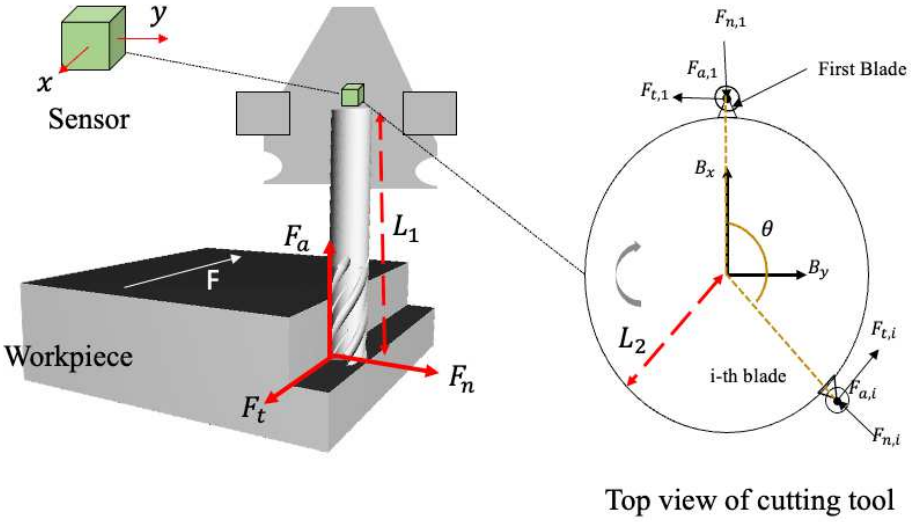


Fig. 2 View of cutting tool

To align the analytical model with the numerical simulations, a generalized discretized bending moment model is utilized, as shown in Fig. 2. Assuming a tool with N_t flutes, the moment contribution of the i^{th} flute in the polar coordinate system is expressed in (3). Where θ is the tool rotation angle, b is the discretized axial depth, $F_{t,i}(b)$ and $F_{n,i}(b)$ are the tangential and radial cutting forces at depth b, and $L_1(b)$ is the corresponding moment arm, and L_2 represents the tool radius. The total bending moments B_x and B_y are calculated as the sum of the bending moments from all flutes, as shown in (4).

$$B_{x,i} = L_1(b)F_{t,i}(b)\,cos(\theta) + L_1(b)F_{n,i}(b)\,sin(\theta) + L_2F_{a,i}\,cos(\theta)$$
$$B_{y,i} = L_1(b)F_{t,i}(b)\,sin(\theta) - L_1(b)F_{n,i}(b)\,cos(\theta) - L_2F_{a,i}\,sin(\theta) \quad (3)$$

$$B_x = \sum_{i=1}^{Nt} B_{x,i}$$
$$B_y = \sum_{i=1}^{Nt} B_{y,i} \quad (4)$$

To verify the derived formulas, the calculated bending moments were compared with numerical simulations. The cutting coefficients are obtained through experiments to calculate the ideal bending moment trajectories (B_x, B_y). These trajectories represent the periodicity and geometric features determined by spindle speed, tool geometry, and cutting parameters. Fig. 3 compares the simulated ideal bending moments with experimental results across different wear stages. The ideal profiles align closely with experimental data during early wear. However, at the end-of-life stage, the geometric variation of the cutting edges causes deviations. This gap between the ideal model and actual observations motivates the use of a conditional diffusion network. By using these ideal moments as physical conditions, the diffusion model is constrained to follow the correct physical frequency, avoiding data that contradict cutting mechanics. The wear-induced dynamics and random noise are then learned from the data through the network.

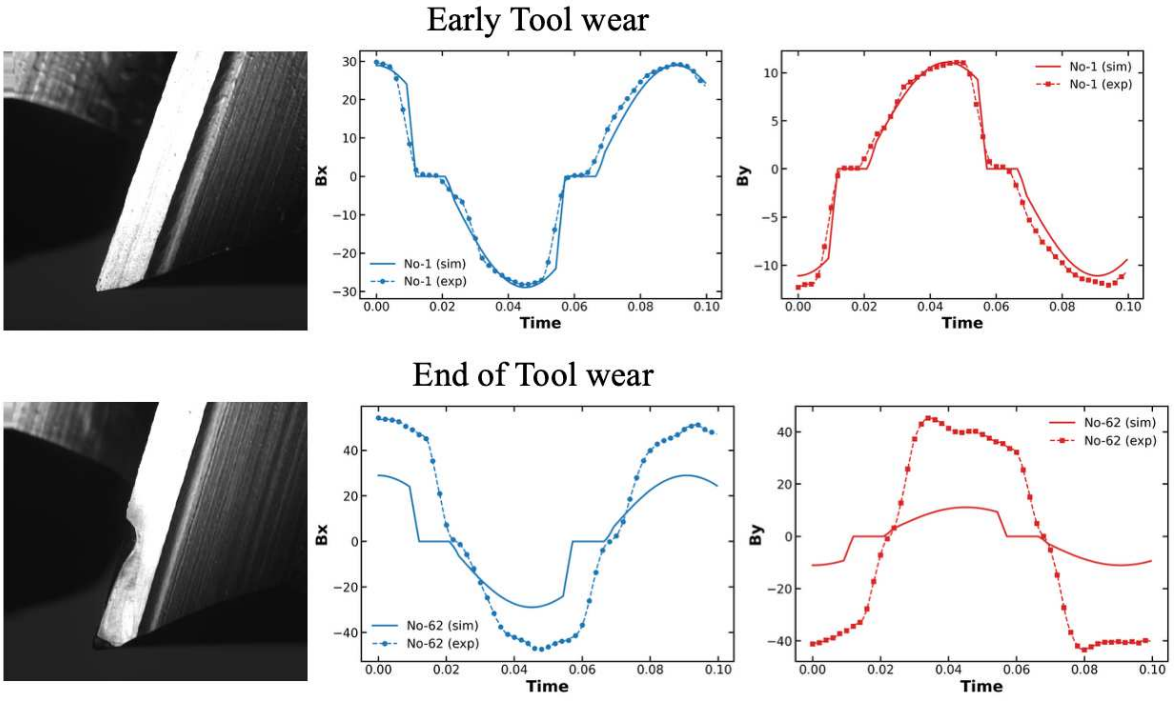


Fig. 3 Bending moment experiment result

2.2. Conditioned Diffusion Network Architecture

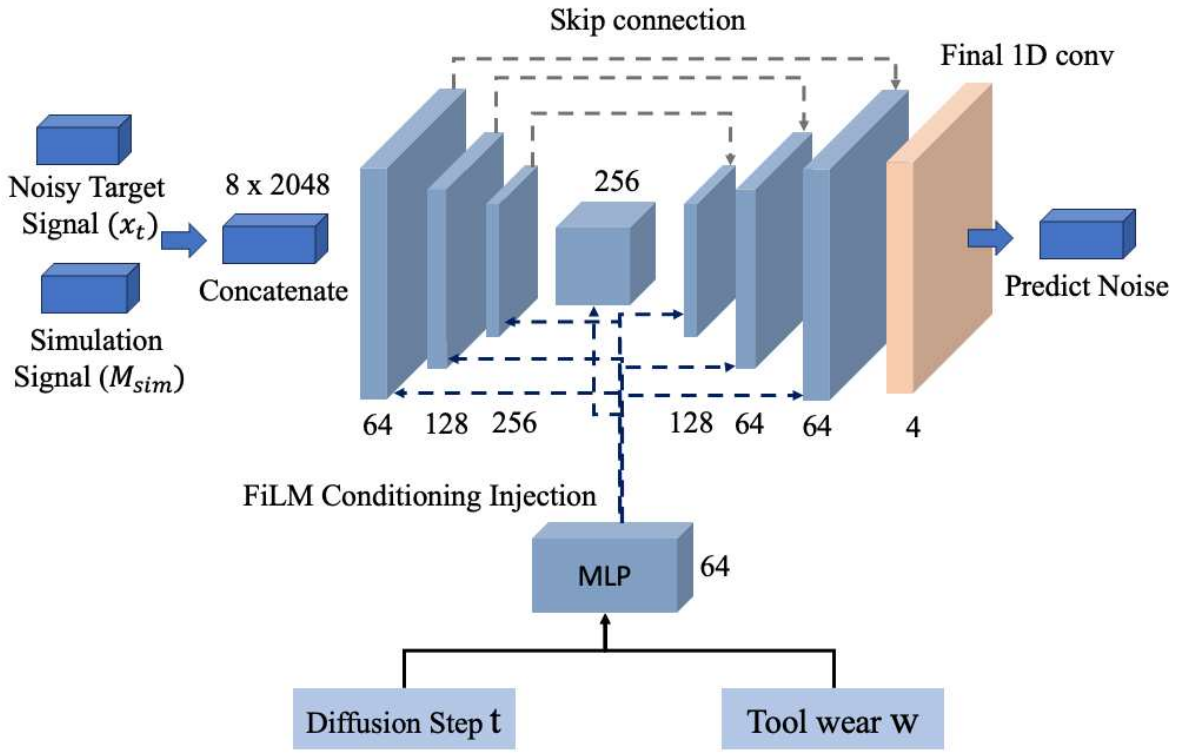


Fig. 4 1D U-Net Diffusion model Architecture

The bending moment signals generated during cutting process are high-frequency time-series data with periodicity tied to tool rotation. The proposed architecture is built upon the Denoising Diffusion Probabilistic Model (DDPM) framework [26], which learns to iteratively denoise a signal from Gaussian noise through a reverse Markov process. To extract features efficiently, this study uses a time-domain 1D-CNN U-Net as backbone for denoising rather than complex frequency-domain transformations. As shown in Fig. 4, the architecture consists of two core components: the 1D-CNN U-Net backbone and the FiLM Block input mechanism. The multi-scale feature extraction of the U-Net enables the network to understand the global periodic profiles determined by spindle speed while preserving signal variations in cutting process. Since these details are easily lost during deep-layer compression, skip connections are used to retain features, preventing overly smoothed or distorted signals.

Actual cutting bending moment includes kinematic periodicity and variations driven by tool wear. To address this problem, the conditioning mechanism is inspired by Stable Diffusion [27]. First, the simulated bending moments (M_{sim}) are concatenated with the noisy signal at the input layer. It is important to emphasize that this integrates physical knowledge directly into the network architecture. The simulated bending moment M_{sim} provides a deterministic structural prior at the input layer, while the FiLM blocks execute dynamic feature-level scaling based on the wear state. This allows the model to identify the ideal period and phase at the start of the generation process, ensuring the model aligns with cutting mechanics.

Second, Feature-wise Linear Modulation (FiLM) is employed to compensate for the non-linear deformations that traditional physical models cannot reflect. The diffusion timestep t

and wear state w are mapped into a latent space and injected into each U-Net block. The conditioning vector e_{cond} is shown in Eq. 5. This vector e_{cond} is then injected into each U-Net block by FiLM layers to modulate the internal features h_{out} as shown in Eq. 6.

Since FiLM dynamically scales and shifts feature channels based on the wear input, the denoising network learns to model wear-related high-frequency components as wear increases. The detailed configuration and layer specifications of the proposed 1D-CNN U-Net are summarized in Table 2. This design provides physical interpretability, enabling the model to transform ideal mathematical signals into realistic wear signals. As shown in Fig. 5, the generated signals (blue) accurately match the experimental data (red).

The proposed U-Net 1D Diffusion model is implemented in PyTorch and trained on Nvidia RTX Pro 6000 GPU. The model is optimized using the AdamW optimizer with an initial learning rate of 2 × 10-4. The training is conducted for up to 1000 epochs with a batch size of 32. The diffusion framework follows a Denoising Diffusion Probabilistic Model (DDPM) schedule with $T = 1000$ diffusion steps. A linear noise schedule is adopted, where the variance β increases from $\beta_{start} = 1 \times 10^{-4}$ to $\beta_{end} = 0.02$.

$$e_{cond} = \mathrm{MLP}([t, w]) \tag{5}$$

$$h_{out} = \gamma_i(e_{cond}) \odot h_i + \beta_i(e_{cond}) \tag{6}$$

Table 2. 1D-CNN U-Net Architecture

Stage	**Output Dimension**	**Details**
Input	8 × 2048	Fusing noisy and physical signal
Initial Embedding	64 × 2048	Kernel = 7
Encoder 1	64 × 1024	Down sampling
Encoder 2	64 × 512	Down sampling
Encoder 3	128 × 256	Down sampling
Bridge	256 × 256	Latent space
Decoder 1	128 × 512	Skip-concatenation (512-128)
Decoder 2	64 × 1024	Skip-concatenation (256-64)
Decoder 3	64 × 2048	Skip-concatenation (128-64)
Output	4 × 2048	Predicting noise

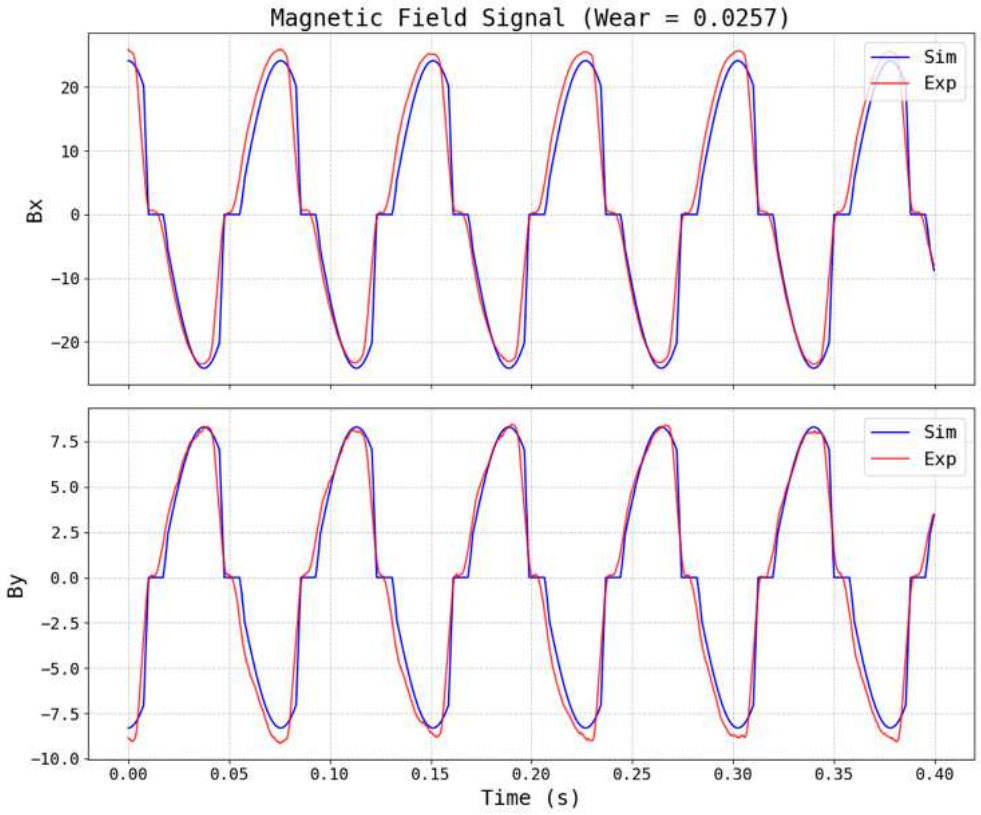


Fig. 5 Comparison between Generated and Experimental Bending Moments

2.3 Downstream Tool Wear Prediction Task

To validate the effectiveness of the generated data, high-fidelity synthetic samples from the diffusion model are mixed with a small amount of real sensor data. This integration constructs an augmented training set designed to enhance the generalization capability of the downstream model. The downstream prediction model utilized a standard 1D-CNN architecture. This network consists of three 1D convolutional layers for local feature extraction, with each followed by Batch Normalization, a ReLU activation function, and a Max Pooling layer. The extracted features are processed through Global Average Pooling and passed to a fully connected network with two hidden layers, outputting a tool wear prediction.

3. Experiment

The cutting experiments in this study were conducted on a five-axis CNC milling machine. To measure the variation in bending moments during the cutting process, a SPIKE intelligent tool holder (manufactured by Pro-Micron, Germany) was utilized for data acquisition. This research applied its high sampling rate to capture real-time bending moment signals, B_x and B_y , in the X and Y directions. Additionally, S45C medium carbon steel, which is widely used in industrial applications, was selected as the workpiece material

To obtain tool wear measurements, this study utilized a custom-designed optical image acquisition system to evaluate the average wear of each cutting edge. As shown in the upper

part of Fig. 6, the tool is captured after the completion of each cutting pass to record the tool wear values.

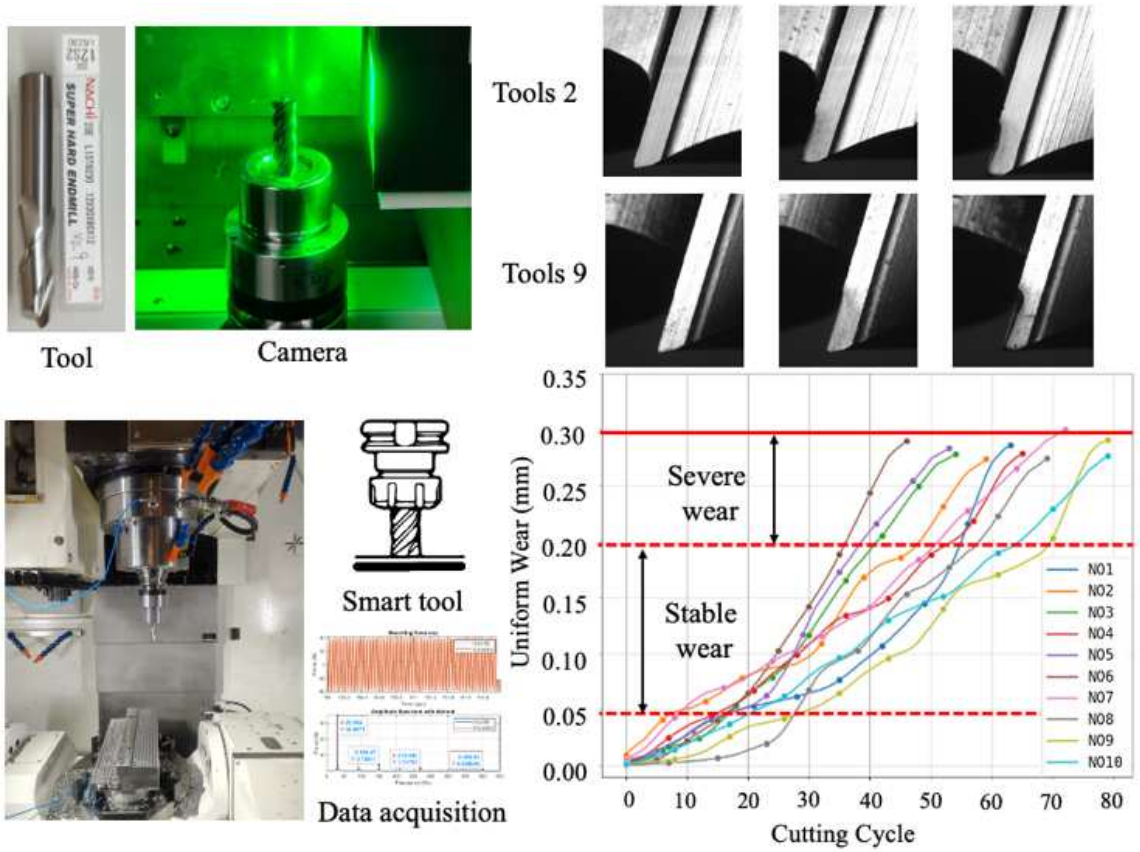


Fig. 6 Cutting experiment platform

Table 3. Experiment details

Equipment name	**Specification**	**Functional notes**
Five-axis CNC machine	HEIDENHAIN TNC-640 Five-axis CNC machine	Milling process.
Smart Tool	Pro-micron spike tool	Acquisition of cutting bending moment.
Workpiece	Material: Carbon Steel: S45C	Workpiece material
Capture tool platform	Measuring the cutting tool	Use to measurement of the uniform VB on cutting tool

To evaluate the generalization capability of the proposed framework, ten sets of machining conditions with different spindle speeds, feed rates, and depths of cut were designed. The detailed cutting parameters are shown in Table 4. For each of the ten tools, experiments were conducted from the initial state until the flank wear reached 0.3 mm. As shown in the wear curves in the lower right of Fig. 6, all tools underwent three stages: initial wear, stable wear, and accelerated wear. This dataset not only enables cross-domain generalization testing but also serves to validate that combining a limited amount of real cutting data with the high-

fidelity bending moment signals generated by this framework is sufficient to train a model for tool wear prediction.

Table 4. Experiment process condition setting

Cutting tools	Spindle speed (rpm)	Feed (mm)	Width (mm)	Depth (mm)	Direction
Tool 1	660	85	10	2	Up
Tool 2	726	93	10	2	Down
Tool 3	792	102	10	2	Up
Tool 4	660	85	10	2	Down
Tool 5	726	85	10	2	Down
Tool 6	792	85	10	2	Up
Tool 7	660	85	8	2	Down
Tool 8	726	93	8	3	Up
Tool 9	792	102	8	3	Up
Tool 10	660	89	10	2.5	Down

A cross-condition validation strategy was implemented for the dataset. Tools 1–7 and Tool 9 were used to construct the training and validation sets. Tools 8 and 10 were reserved entirely to serve as a test set representing unseen machining parameter conditions. Unlike a random split, this setup was designed to test how the framework handles parameter-shift scenarios. Specifically, while some individual parameter values exist in the training set, Tool 8 and Tool 10 operate under completely unseen combinations of spindle speeds, radial widths, and axial depths, alongside a newly introduced 2.5 mm cutting depth for Tool 10. This rigorous partitioning strategy ensures that the test data remains completely unseen during the generative model's training phase. This setup evaluates the framework's generation quality and predictive performance in practical cross-domain scenarios. Regarding the data sampling strategy, random sampling was performed on the steady-state physical characteristics during the cutting process. In these experiments, the total processing time for a single milling process is approximately 220 seconds. Since the machine maintains a stable state during cutting, the generated bending moment signals show periodicity and repeatability. Accordingly, 10 data segments of 0.4 seconds each were randomly extracted from the stable cutting region of each pass to capture the corresponding wear feature. The total number of samples per tool is summarized in Table 5.

Table 5. Dataset composition and partitioning details

Cutting Tools	Dataset Role	Number of Cutting	Total Samples
Tool 1	Training	64	640
Tool 2	Training	60	600
Tool 3	Training	55	550
Tool 4	Training	66	660

Tool 5	Training	54	540
Tool 6	Training	47	470
Tool 7	Training	73	730
Tool 8	Testing	70	700
Tool 9	Training	80	800
Tool 10	Testing	80	800

297

298 A quantitative analysis is conducted from the dimensions of synthetic signal fidelity and
299 wear prediction effectiveness to evaluate the performance of the proposed framework.
300 Regarding the signal fidelity evaluation, the synthetic bending moments generated by the
301 diffusion model are compared with experimental data in both the time and frequency domains.
302 Fast Fourier Transform (FFT) is utilized to analyze the spectral characteristics of the
303 generated signals, verifying whether the physical harmonics determined by the spindle speed
304 and tooth passing frequency are preserved. Furthermore, the t-Distributed Stochastic
305 Neighbor Embedding (t-SNE) algorithm is applied to visualize high-dimensional features in
306 a 2D space. This step verifies whether the distribution of wear-related features in the synthetic
307 data aligns with real experimental observations and ensures that the generated data does not
308 induce negative transfer in downstream tasks.
309 For the evaluation of wear prediction effectiveness, only a minimal amount of real
310 samples is randomly selected from the training tools to serve as the baseline. The core
311 objective of the diffusion model is to utilize this extremely small-scale real dataset to generate
312 a large volume of high-fidelity synthetic bending moments. This process augments the
313 training set and effectively fills the gaps within the feature space. Finally, Root Mean Square
314 Error (RMSE) is employed as the quantitative metric to assess the model's predictive
315 performance on the test set.
316
317
318

319 **4. Results and Discussion**

320 4.1 Fidelity Analysis of Generated Cutting Bending Moment Signals

321 To evaluate the quality of the generated bending moment signals, a data-driven model,
322 WGAN, was also implemented for comparison. Following a similar training protocol, data
323 from Tools 1–7 and Tool 9 were utilized as the training set. However, no separate test set
324 was defined for the WGAN, as its primary objective was to learn feature distributions through
325 adversarial learning.
326 Additionally, a physics-guided WGAN (PG-WGAN) was developed, incorporating
327 physical cutting formulas and tool wear values as conditional inputs. Processing high-
328 frequency periodic signals is challenging for conventional GAN models lacking physical
329 constraints. As a result, these models often suffer from mode collapse, a phenomenon
330 illustrated in Fig. 7. Experimental observations indicate that even when the adversarial

training converges, the WGAN-generated signals do not conform to the physical laws of machining dynamics. Generative models utilizing stochastic noise processes inherently tend to introduce artifacts into the primary signal components. In the case of WGAN, the model synthesizes a probable representation based on the features learned during adversarial training. However, without the guidance of physical laws, this process often leads to the superimposition of significant noise over the periodic harmonics. PG-WGAN improves the alignment of frequency components, as shown in Fig. 9. However, its overall effectiveness remains limited in terms of amplitude accuracy.

Furthermore, the variation in signal amplitude is limited by the training data. Consequently, in the final stages of tool wear, the generated magnitudes fail to align with experimental observations. This deviation is further manifested in the Polar Trajectory Plots (Fig. 8), where the synthesized phase angles and magnitudes exhibit significant departures from the experimental results during the late-stage wear.

In contrast, the proposed bending moment diffusion model utilizes multi-dimensional feature validation. Fig. 11 compares the synthesized and experimental signals (B_x, B_y) for Tool 10 at both the initial and severe wear stages. Although a temporal phase shift exists in the time domain, the frequency-domain analysis confirms that the model accurately reconstructs the physical harmonics associated with spindle speed and cutting edge dynamics. As tool geometry degrades with increasing wear, the generated outputs maintain high physical consistency with the experimental data. This fidelity is further visualized in the Polar Trajectory Plots (Fig. 12), where the synthesized paths exhibit high overlap with the ground truth. As summarized in Table 6, the quantitative metrics across different wear stages further substantiate these observations. The Proposed method maintains a low Resultant Moment Deviation (RMD) ranging from 0.19 to 4.86 Nm, with high FFT Spectral Correlation. The WGAN baseline exhibits RMD of 47.53 Nm in the final wear stage and fails to capture the physical laws. Even with the inclusion of physical constraints, the PG-WGAN (Physical-Guided) shows only marginal improvements in spectral correlation and still with amplitude fluctuations. This indicates that the adversarial framework is less effective than the diffusion-based approach in reconciling stochastic generation with rigid physical boundaries.

To evaluate signal generation under multiple unseen tool conditions, an additional 5-fold validation was conducted. In each fold, eight tool conditions were used for training, and the remaining two tool conditions were held out for evaluation. As shown in Table 7, across the ten held-out tool conditions, the average resultant moment deviation was 5.268 Nm, and the average FFT Spectral Correlation was 0.9496 for the resultant moment. These results show that the proposed model preserves the main frequency-domain characteristics under multiple unseen tool conditions.

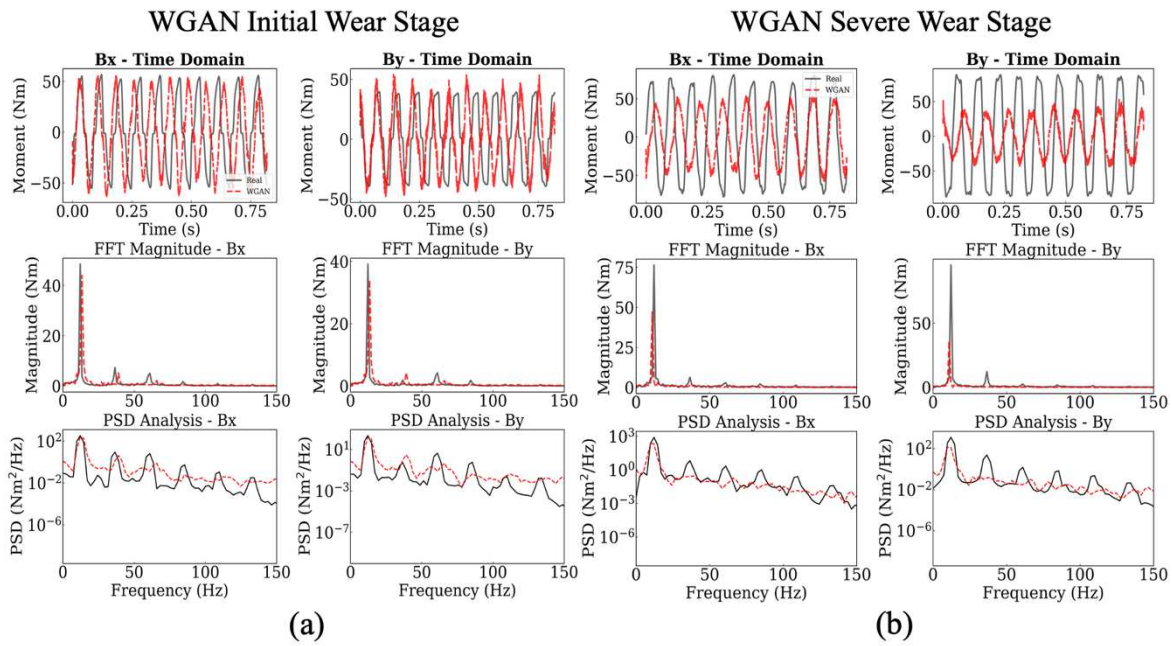


Fig. 7：WGAN Bending Moment Generate Result (a) Initial Wear Stage (b) Severe Wear Stage

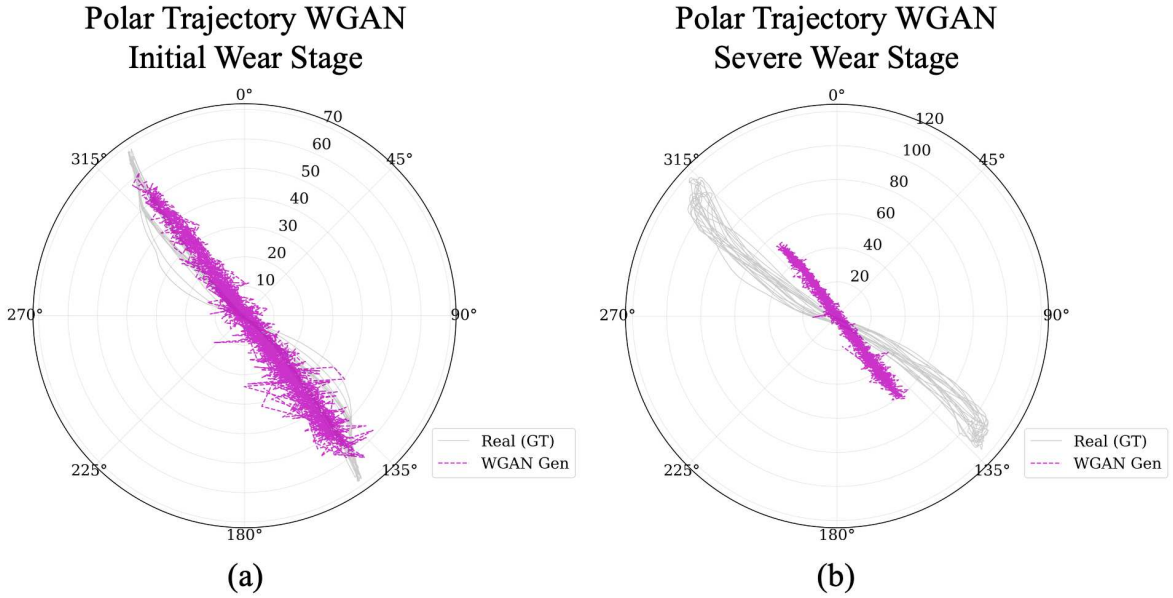


Fig. 8 Polar Trajectory of WGAN Cutting Bending moment (a) Initial Wear Stage (b) Severe Wear Stage

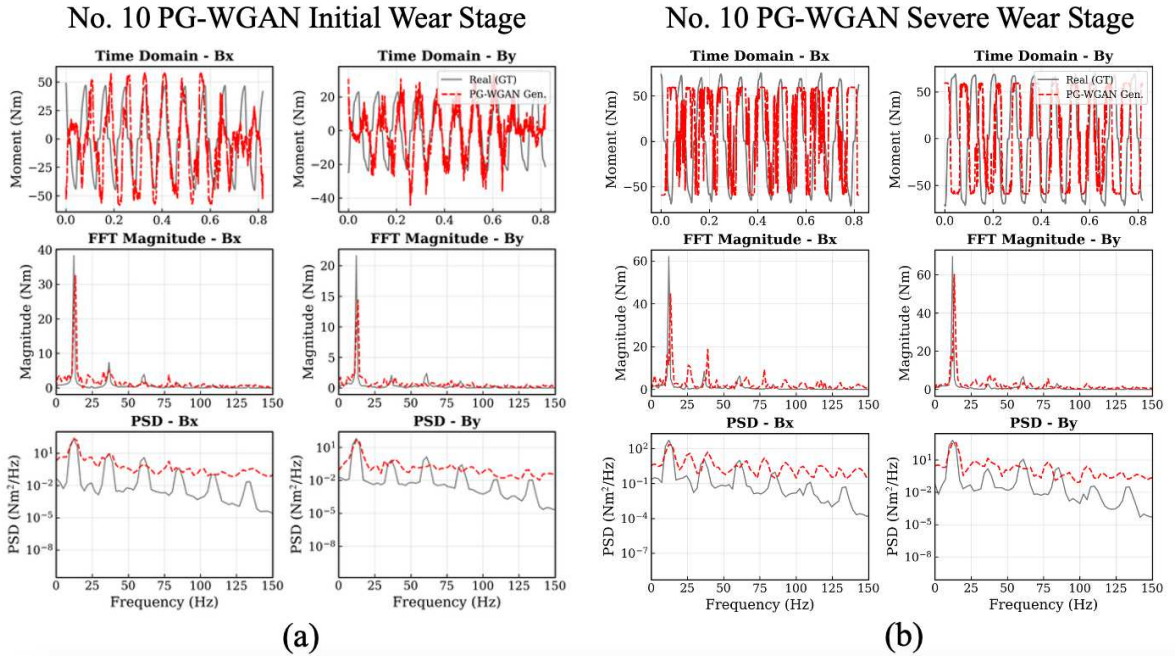


Fig. 9：PG-WGAN Bending Moment Generate Result (a) Initial Wear Stage (b) Severe Wear Stage

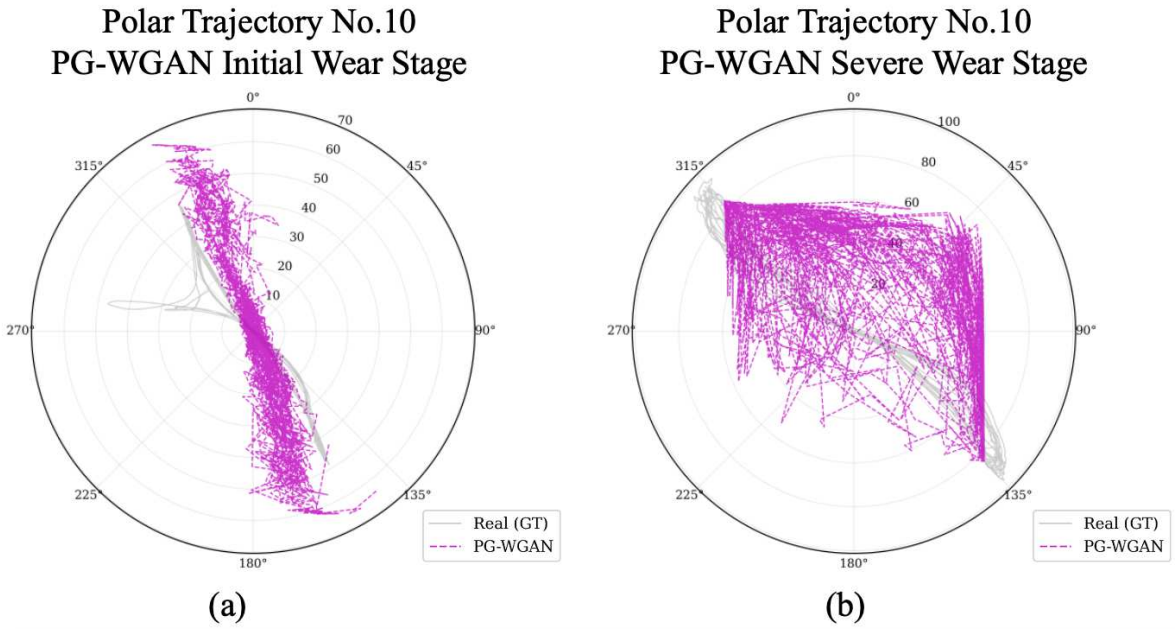


Fig. 10：Polar Trajectory of PG-WGAN Cutting Bending moment (a) Initial Wear Stage (b) Severe Wear Stage

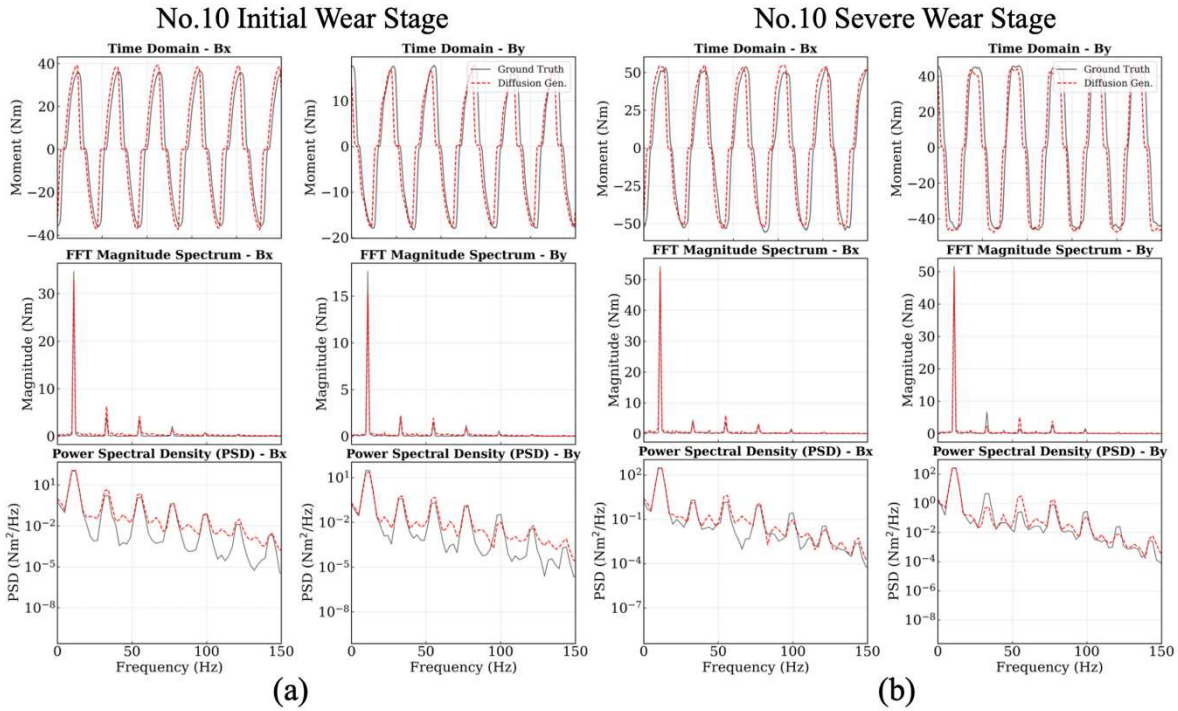


Fig. 11 Tool 10 Severe Wear Stage (a) Initial Wear Stage (b) Severe Wear Stage

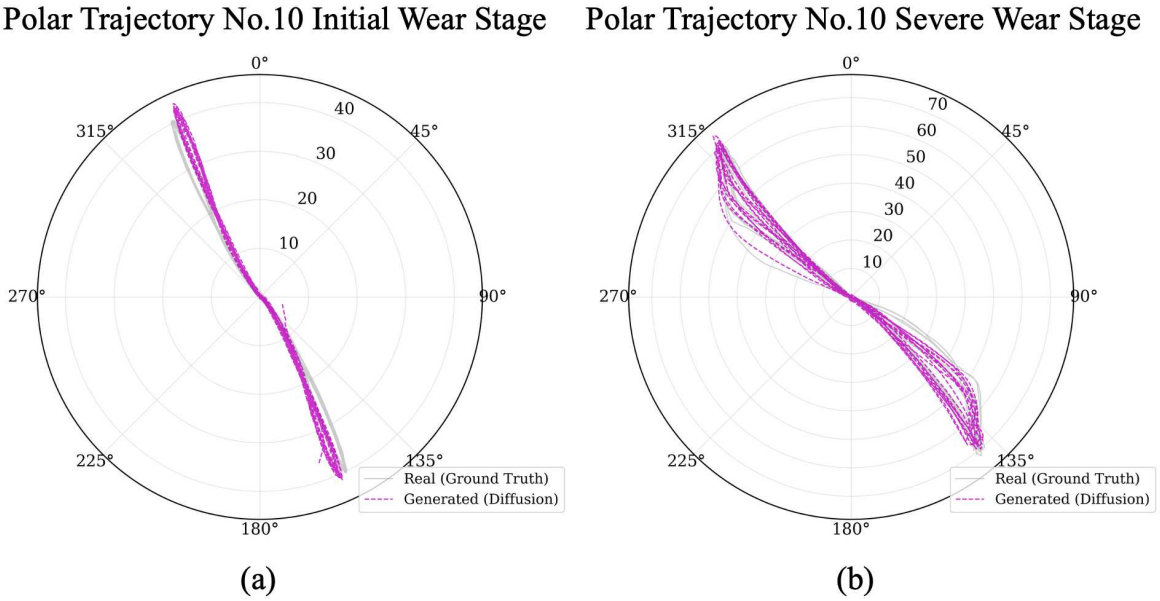


Fig. 12 Polar Trajectory of Tool 10 Cutting Bending moment (a) Initial Wear Stage (b) Severe Wear Stage

401

402

403 **Table 6.** Performance metrics of generative models across different wear stages.

Method	Condition / Stage	Resultant Moment (Nm)	FFT Spectral Corr.
WGAN (Data Driven)	Wear Initial	3.7295	0.2194
	Wear Mid	26.0644	0.1775
	Wear Final	**47.5338**	0.3811
PG-WGAN (Physical-Guide)	NO8 Initial	15.512	0.6925
	NO8 Mid	27.1931	0.6440
	NO8 Final	16.5943	0.7088
	NO10 Initial	12.8420	0.8315
	NO10 Mid	31.5532	0.9501
	NO10 Final	15.5124	0.8524
Proposed	NO8 Initial	0.3513	0.9945
	NO8 Mid	3.0458	0.9953
	NO8 Final	0.1935	0.9948
	NO10 Initial	2.3440	0.9933
	NO10 Mid	4.8619	0.9964
	NO10 Final	2.9527	0.9972

404

405 **Table 7.** Cross-condition 5-fold validations results.

Tool	Fold	Resultant Moment (Nm)	FFT Spectral Correlation
No1	1	3.8596	0.9453
No2	1	4.8519	0.9364
No3	2	4.4670	0.9698
No4	2	3.8028	0.9691
No5	3	4.9474	0.9461

No6	3	5.0642	0.9631
No7	4	4.0959	0.9163
No8	4	7.6891	0.9585
No9	5	7.99	0.9391
No10	5	5.922	0.9531
Average	-	5.2689	0.9496

406

407 **Table 8.** Performance evaluation of architectural components through ablation analysis.

Method	MAE (Nm)	FFT Spectral Corr.	Description
Full Method	~2.289	0.99	Complete architecture
Without Physical	~32.5	0.28	Purely data-driven diffusion
Without FiLM	~64.5	0.69	Wear Value for Feature

408

409 4.2 Data distribution diversity

410 The preceding analysis of physical fidelity demonstrates the model's capability to
411 synthesize bending moments across various wear stages. To evaluate the diversity of the
412 generated data, the t-SNE (t-distributed Stochastic Neighbor Embedding) algorithm is
413 employed for feature visualization. The objective is to determine its efficacy in augmenting
414 the data distribution and addressing the sparsity within experimental datasets. As a
415 nonlinear dimensionality reduction technique, t-SNE is utilized to project high-dimensional
416 features onto a two-dimensional plane. By mapping both experimental and diffusion-
417 generated signals into this 2D space, the underlying distribution characteristics and data
418 clusters can be observed.

419 However, calculating distances directly in the time-domain space is problematic due to
420 minor phase shifts within the signals. Such temporal misalignments could distort the
421 resulting manifold distribution. To address this issue, a Fast Fourier Transform (FFT) is
422 applied to extract the 0–500 Hz frequency band, which contains the cutting energy and
423 deterministic physical harmonics. These frequency-domain components serve as the feature
424 vectors for the t-SNE analysis, ensuring that the visualization reflects the true physical
425 similarities between signals.

426 Fig. 13, Fig. 14 and Fig. 15 illustrate the t-SNE feature distributions for the unseen test
427 sets (Tool 8 and Tool 10) and the training set (Tool 1). Conducting analysis directly on
428 unseen conditions represents a test of the generative model's cross-condition
429 generalization. As observed in the plots, the synthetic samples generated by the diffusion
430 model exhibit distribution diversity. These samples effectively interpolate and populate
431 the gaps between experimental data clusters.

As observed in Fig. 14, minor translations in absolute coordinates occur between generated and experimental features due to unseen cutting parameters. Nevertheless, the overall distribution of synthetic samples remains parallel to the ground truth. The gradient distribution plots on the right are color-coded by wear values. These plots reveal that the trajectories for both Tool 8 and Tool 10 strictly follow the physical evolution of tool wear from initial to severe stages. To provide a quantitative validation of the distribution similarity, the Maximum Mean Discrepancy (MMD) was calculated across different wear stages, as summarized in Table 9. The results show that the MMD value for the initial and middle stages remain at low levels. Although a higher MMD is observed in the final stage, this is attributed to the stochastic vibrations of severe tool wear. These combined qualitative and quantitative results demonstrate that the proposed model architecture effectively captures the physical frequency characteristics of machining dynamics and the evolutionary trends associated with tool wear progression.

Table 9. Quantitative evaluation of data distribution using MMD.

Case	Initial Stage	Mid Stage	Final Stage
No1	0.10	0.285	0.38
No8	0.125	0.143	2.0
No10	0.182	0.105	0.40

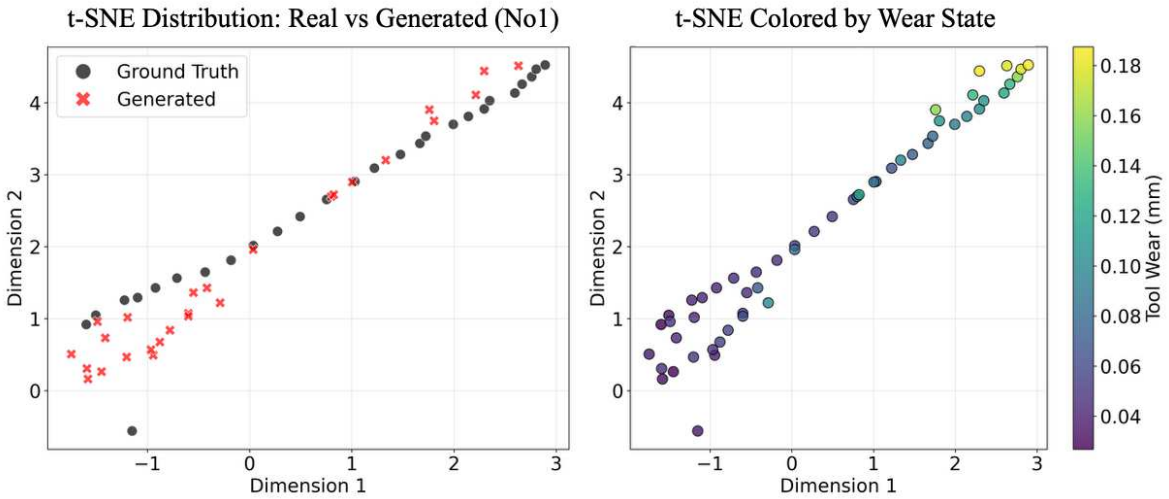


Fig. 13：Tool 1 t-SNE Result

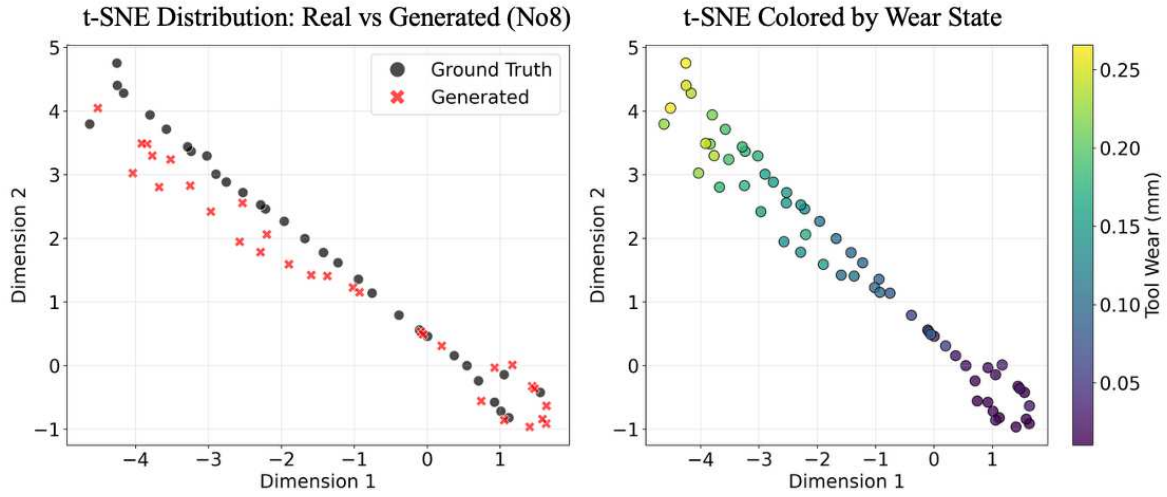


Fig. 14：Tool 8 t-SNE Result

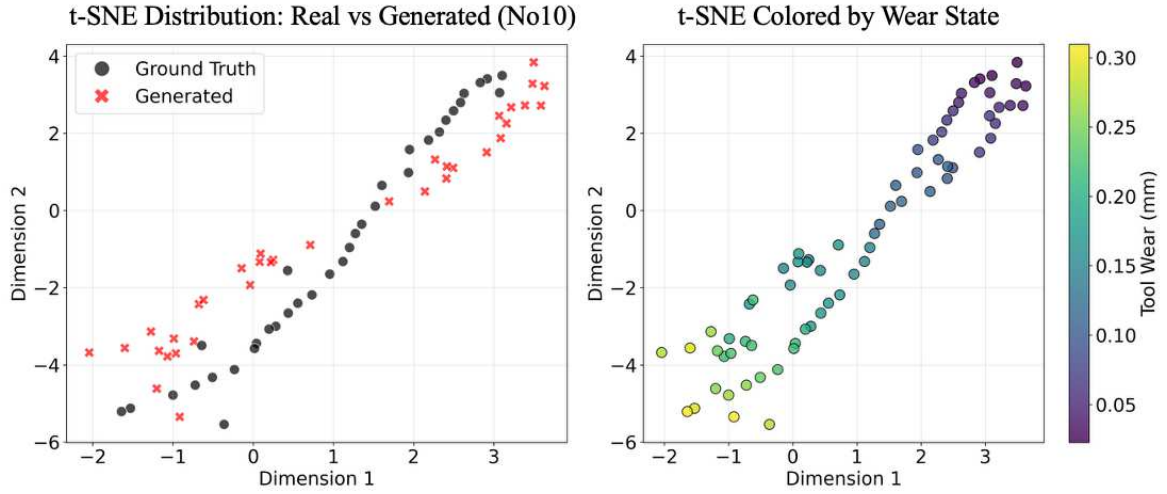


Fig. 15：Tool 10 t-SNE Result

4.3 Tool wear prediction results across diverse cutting conditions

As established in the previous section, the proposed bending moment diffusion model is capable of synthesizing time-domain data for various wear states under different cutting conditions. To evaluate whether these generated signals can be effectively applied to downstream tasks, the 1D-CNN tool wear prediction model was trained using a limited dataset. A baseline training set was constructed by randomly sampling only 5% of the real-world data from Tools 1–7 and Tool 9. At the same time, the diffusion model was utilized to generate 300 synthetic signals for the unseen cutting parameters of Tool 8 and Tool 10. These synthetic samples covered a uniform wear range between 0 and 0.3 mm. These Synthetic samples were incorporated into the 5% real-world training set. Throughout the training process, the 1D-CNN model remained entirely unexposed to any actual sensor signals from Tool 8 or Tool 10, ensuring a cross-condition validation. The resulting Root Mean Square Error (RMSE) metrics for these tests are summarized in Table 10.

Experimental results indicate that the Baseline model suffers from severe prediction offsets when encountering the unseen conditions of Tool 8 and Tool 10. Stemming from the

limited 5% training data as shown in Fig. 16, the available samples were insufficient to encompass the comprehensive feature characteristics required for accurate cross-condition prediction. The average RMSE for these unseen cases reaches as high as 0.04 mm. To increase the diversity of the cutting data, the WGAN method and proposed bending moment generation model were imported. Fig. 17 and Fig. 19 illustrate the tool wear prediction trajectories under the 5% data constraint, comparing the proposed method with the WGAN augmentation strategy. The Baseline model (blue line) exhibits significant trajectory deviation during the mid-to-late stages of tool wear due to insufficient training data. In contrast, the proposed strategy in Fig. 17 shows that the curve maintains close alignment with the Ground Truth even during the severe wear stages. However, as depicted in Fig. 19, the WGAN generation strategy fails to provide accurate guidance for the downstream task. Resulting from a deficiency in physical constraints, the synthetic samples mislead the model's direction. This results in a loss of prediction accuracy and manifests as the negative transfer previously discussed.

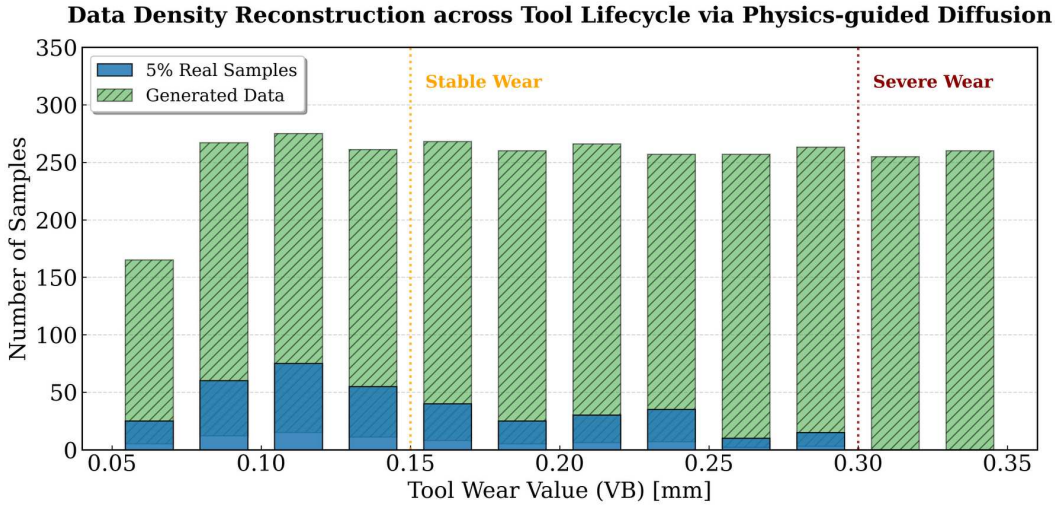


Fig. 16：Data density reconstruction across the tool lifecycle by physics-guided diffusion augmentation.

In addition to the 5% data experiments, tests were conducted using the full 100% training dataset to evaluate whether synthetic augmentation still impacts performance when real data is abundant. As shown in the results, when augmenting the 100% dataset, the prediction trajectory of the WGAN-augmented model aligns more closely with that of the proposed method (as shown in Fig. 20). This suggests that with sufficient training data, the downstream model could maintain convergence and prediction accuracy even when unconstrained synthetic data is introduced. However, in the final stages of severe tool wear, the WGAN-augmented results still exhibit relatively poorer performance.

Table 10. Quantitative comparison of tool wear prediction performance under different data augmentation methods

Experiment Setup	Method	RMSE	MAE	R^2
Data Scarcity (5% Real)	Baseline (Real Only)	0.0440	0.0516	0.4110
	WGAN (Real + WGAN)	0.0627	0.0379	0.6921
	PG-WGAN (Real + Physical WGAN)	0.0517	0.0460	0.6299
	Propose (Real + Diffusion)	**0.0155**	**0.0201**	**0.9321**
Full Data (100% Real)	Baseline (Real Only)	0.0265	0.0147	0.95735
	WGAN (Real + WGAN)	0.0350	0.0269	0.87425
	PG-WGAN (Real + Physical WGAN)	0.0340	0.0160	0.94855
	Propose (Real + Diffusion)	**0.0244**	**0.0117**	**0.9713**

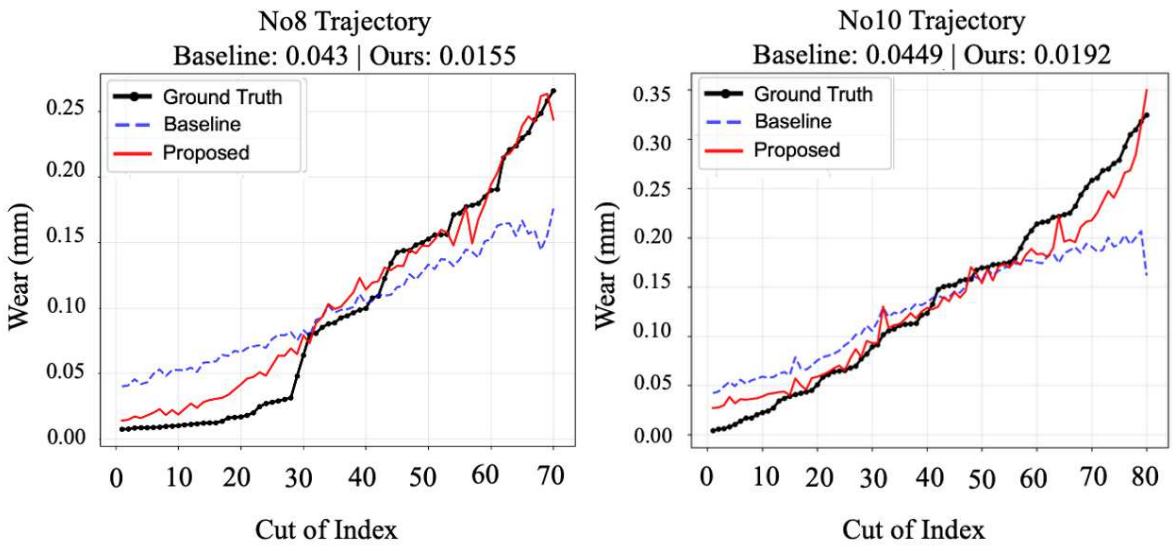
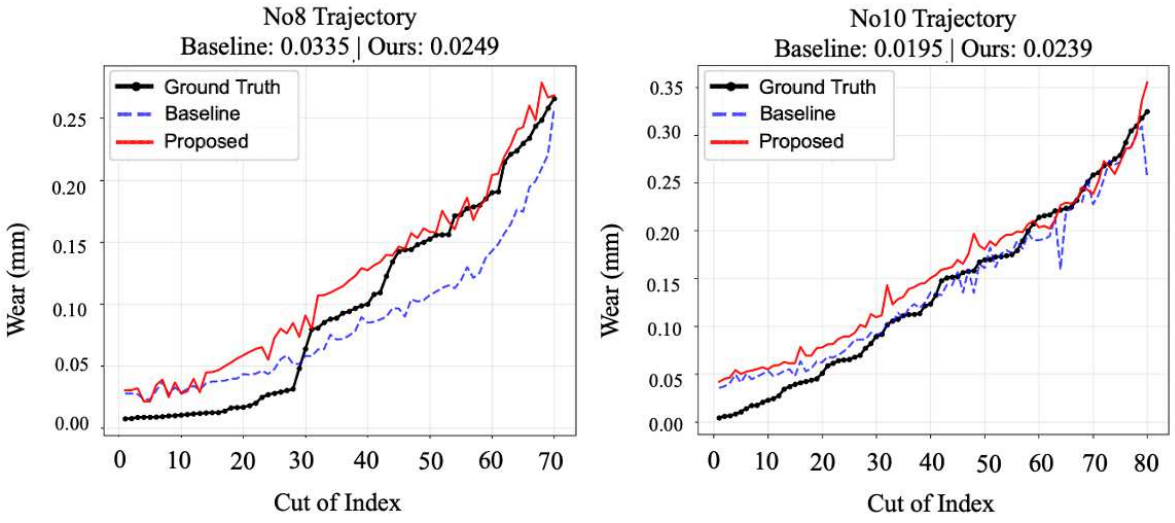
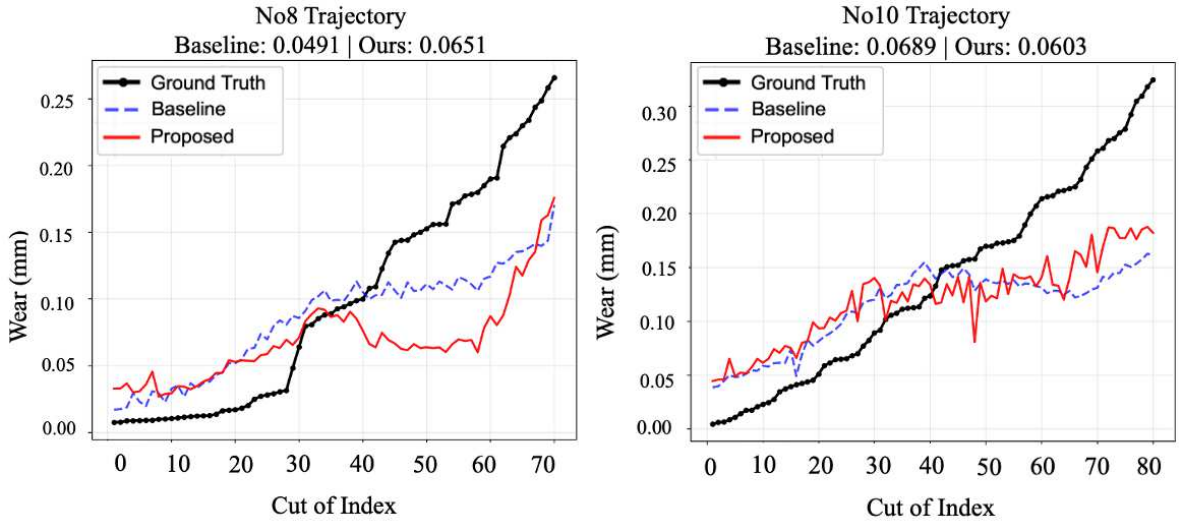


Fig. 17：5% Real with propose method



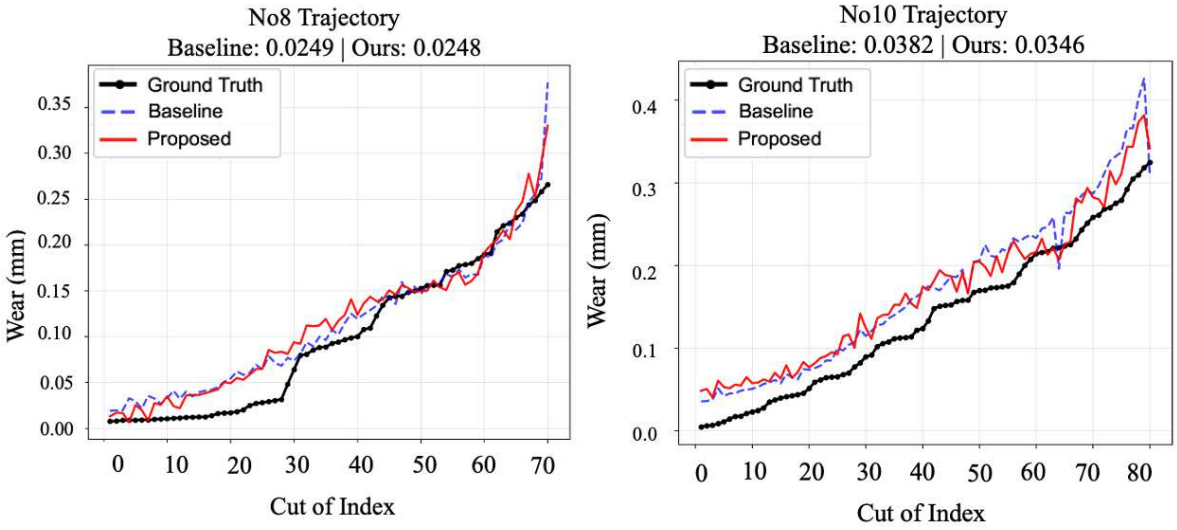
503

504 Fig. 18：100% Real with propose method



505

506 Fig. 19：5% Real with WGAN method



507

508 Fig. 20：100% Real with WGAN method

5. Conclusion

This study proposes a physics-guided diffusion model for the synthesis of cutting bending moments. The system integrates an idealized cutting bending model with a 1D-CNN U-Net backbone to generate high-fidelity signals across various tool wear stages. By utilizing the FiLM Block, the model effectively modulates the denoising process based on specific input conditions to produce corresponding bending moments. These synthesized data were subsequently employed to train a downstream 1D-CNN model for tool wear prediction. Experimental results demonstrate that conventional data-driven generative models struggle to maintain physical features when processing high-frequency signals, often leading to severe negative transfer. In contrast, the proposed method enhances predictive performance, achieving a 60.45% reduction in average RMSE for unseen cutting conditions. Finally, the bending moment generation model developed in this research effectively addresses the limitations of conventional experimental approaches, eliminating the need for exhaustive full-life tool wear tests whenever operating conditions change for tools with the same number of flutes.

Declarations

- **Funding**: This research is supported by the National Science and Technology Council (NSTC), R.O.C., under the contracts 114-2218-E-002-007 and 114-2218-E-002-027.
- **Competing interests**: The authors declare no competing interests.
- **Author contribution**: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yu-Sheng Lin and Jhuang-Jian Ping. The first draft of the manuscript was written by Yu-Sheng Lin. Meng-Shiun Tsai and Yu-Sheng Lin commented on previous versions of the manuscript. All authors read and approved the final manuscript.
- **Data and Code Availability**: The data and code that support the findings of this study are available from the corresponding author upon reasonable request. A complete repository containing the core implementation of the physics-guided diffusion model will be open-sourced at https://github.com/kmolLin/Cutting-Bending-Moment-Integrated-Diffusion-Model in the near future.

References

1. Yue C, Gao H, Liu X, Liang SY (2018) Part Functionality Alterations Induced by Changes of Surface Integrity in Metal Milling Process: A Review. Appl Sci 8:2550. https://doi.org/10.3390/app8122550

2. Li B (2012) A review of tool wear estimation using theoretical analysis and numerical simulation technologies. Int J Refract Met Hard Mater 35:143–151. https://doi.org/10.1016/j.ijrmhm.2012.05.006

3. Jurkovic J, Korosec M, Kopac J (2005) New approach in tool wear measuring technique using CCD vision system. Int J Mach Tools Manuf 45:1023–1030. https://doi.org/10.1016/j.ijmachtools.2004.11.030

4. Serin G, Sener B, Ozbayoglu AM, Unver HO (2020) Review of tool condition monitoring in machining and opportunities for deep learning. Int J Adv Manuf Technol 109:953–974. https://doi.org/10.1007/s00170-020-05449-w

5. Mohanraj T, Kirubakaran ES, Madheswaran DK, Naren ML, Dharshan SP, Ibrahim M (2024) Review of advances in tool condition monitoring techniques in the milling process. Meas Sci Technol 35:092002. https://doi.org/10.1088/1361-6501/ad519b

6. Kumar SG, Singh BK, Kumar RS, Haldorai A (2023) Digital Twin Framework for Lathe Tool Condition Monitoring in Machining of Aluminium 5052. Def Sci J 73:341–350. https://doi.org/10.14429/dsj.73.18650

7. Manwar A, Varghese A, Bagri S, Joshi SS (2025) Online tool condition monitoring in micromilling using LSTM. J Intell Manuf 36:935–955. https://doi.org/10.1007/s10845-023-02273-3

8. Zhou Y, Sun W, Ye C, Peng B, Fang X, Lin C, Wang G, Kumar A, Sun W (2023) Time-frequency Representation-enhanced Transfer Learning for Tool Condition Monitoring during milling of Inconel 718. Eksploat Niezawodn-Maint Reliab 25:165926. https://doi.org/10.17531/ein/165926

9. Xu X, Wang J, Zhong B, Ming W, Chen M (2021) Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism. Measurement 177:109254. https://doi.org/10.1016/j.measurement.2021.109254

10. Huang Z, Shao J, Zhu J, Zhang W, Li X (2024) Tool wear condition monitoring across machining processes based on feature transfer by deep adversarial domain confusion network. J Intell Manuf 35:1079–1105. https://doi.org/10.1007/s10845-023-02088-2

11. Li C, Li S, Feng Y, Gryllias K, Gu F, Pecht M (2024) Small data challenges for intelligent prognostics and health management: a review. Artif Intell Rev 57:214. https://doi.org/10.1007/s10462-024-10820-4

12. Shao S, Wang P, Yan R (2019) Generative adversarial networks for data augmentation in machine fault diagnosis. Comput Ind 106:85–93. https://doi.org/10.1016/j.compind.2019.01.001

13. Shah M, Vakharia V, Chaudhari R, Vora J, Pimenov DYu, Giasin K (2022) Tool wear prediction in face milling of stainless steel using singular generative adversarial network and LSTM deep learning models. Int J Adv Manuf Technol 121:723–736. https://doi.org/10.1007/s00170-022-09356-0

14. Liu H, Zhao H, Wang J, Yuan S, Feng W (2022) LSTM-GAN-AE: A Promising Approach for Fault Diagnosis in Machine Health Monitoring. IEEE Trans Instrum Meas 71:1–13. https://doi.org/10.1109/TIM.2021.3135328

15. He J, Xu Y, Pan Y, Wang Y (2024) Adaptive weighted generative adversarial network with attention mechanism: A transfer data augmentation method for tool wear prediction. Mech Syst Signal Process 212:111288. https://doi.org/10.1016/j.ymssp.2024.111288

16. Croitoru F-A, Hondru V, Ionescu RT, Shah M (2023) Diffusion Models in Vision: A Survey. IEEE Trans Pattern Anal Mach Intell 45:10850–10869. https://doi.org/10.1109/TPAMI.2023.3261988

17. Liu D, Liu Z, Wang B, Song Q, Li L, Jiang A (2025) Challenges of randomness in tool wear with small samples: A physics-informed cross-domain monitoring method. J Manuf Syst 80:694–722. https://doi.org/10.1016/j.jmsy.2025.04.002

18. Chen X, Zhang H, Fu Y, Song Q, Zhang H, Yi J, Zhou Y (2025) Symmetrized dot pattern enhanced cross-attention DDPM method for tool condition monitoring. AIP Adv 15:. https://doi.org/10.1063/5.0300586

19. Zhu Q, Sun B, Zhou Y, Sun W, Xiang J (2021) Sample Augmentation for Intelligent Milling Tool Wear Condition Monitoring Using Numerical Simulation and Generative Adversarial Network. IEEE Trans Instrum Meas 70:1–10. https://doi.org/10.1109/TIM.2021.3077995

20. Gonzalo O, Beristain J, Jauregi H, Sanz C (2010) A method for the identification of the specific force coefficients for mechanistic milling simulation. Int J Mach Tools Manuf 50:765–774. https://doi.org/10.1016/j.ijmachtools.2010.05.009

21. Yellowley I (1985) Observations on the mean values of forces, torque and specific power in the peripheral milling process. Int J Mach Tool Des Res 25:337–346. https://doi.org/10.1016/0020-7357(85)90034-4

22. Cai S, Mao Z, Wang Z, Yin M, Karniadakis GE (2021) Physics-informed neural networks (PINNs) for fluid mechanics: A review

23. Hibbeler RC by Russell C. Hibbeler Mechanics of Materials (8th Edition) (Alternative eText Formats)(text only)8th (Eighth) edition[Hardcover]2010

24. Perez E, Strub F, Vries H de, Dumoulin V, Courville A (2017) FiLM: Visual Reasoning with a General Conditioning Layer

25. Birnbaum S, Kuleshov V, Enam Z, Koh PW, Ermon S (2021) Temporal FiLM: Capturing Long-Range Sequence Dependencies with Feature-Wise Modulations

26. Ho J, Jain A, Abbeel P (2020) Denoising Diffusion Probabilistic Models

27. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2022) High-Resolution Image Synthesis with Latent Diffusion Models