

Generating Metamorphic Test Cases using 2-qubit Quantum Gate Combinations

Woo-Yeob Hyun*, Dongsu Kang**

*Professor, Dept. of AI / System Science, Korea Army Academy at Yeongcheon, Yeongcheon, Korea

**Professor, Dept. of Cyber and Computer Engineering, Korea National Defense University, Nonsan, Korea

[Abstract]

In quantum computing, the output probability distribution may vary with each execution of the same circuit due to its sensitivity to noise. In this study, 2-qubit quantum gate combinations that are expected to produce identical output distributions under ideal conditions are designed as metamorphic test cases, and Depolarizing, Amplitude Damping, and Phase Damping noise models are applied to analyze the effect of each noise type on the output probability distribution. Experimental results show that in all noise models, when the noise probability reached 0.01, the p-value dropped below the significance level, causing the collapse of the defined metamorphic relation. It was also quantitatively confirmed through KL-Divergence and Cosine Similarity that the divergence from the expected output distribution accumulates with increasing noise strength. When applying metamorphic testing to quantum gate combinations, quantum gate testing can be performed by verifying the feasibility of metamorphic relation judgment and quantitatively tracking the process using the three proposed metrics. The results of this study can be applied to quantum computing reliability evaluation and noise-tolerant circuit design, and present the potential as a versatile quantum software testing methodology.

▶ **Key words:** Quantum Computing, Noise Model, Quantum Gate Combination, Metamorphic Testing, Metamorphic Relation

[요 약]

양자 컴퓨팅은 노이즈 민감성으로 인해 동일한 회로를 반복 실행하더라도 출력 확률 분포가 매번 달라질 수 있다. 본 연구는 이상적인 환경에서 동일한 출력 분포를 예상할 수 있는 2-큐비트 양자 게이트 조합을 메타모픽 테스트 케이스로 설계하고, Depolarizing, Amplitude Damping, Phase Damping 노이즈 모델을 적용하여 각 노이즈가 출력 확률 분포에 미치는 영향을 분석하였다. 실험 결과, 모든 노이즈 모델에서 노이즈가 0.01에 도달하자 p-value가 유의수준 이하로 떨어져 정의한 메타모픽 관계가 붕괴되었으며, KL-Divergence와 Cosine Similarity를 통해 예상 출력 분포와의 차이가 노이즈 강도에 따라 누적됨을 정량적으로 확인하였다. 메타모픽 테스트 기법을 양자 게이트 조합에 적용할 때, 제안한 세 가지 지표를 활용하여 메타모픽 관계 판정 가능 여부를 확인하고 그 과정을 정량적으로 추적함으로써 양자 게이트 테스트를 수행할 수 있다. 연구 결과는 양자 컴퓨팅 신뢰성 평가 및 노이즈 내성 회로 설계 등에 활용될 수 있으며, 다양한 양자 소프트웨어 테스트 방법론으로 가능성을 제시한다.

▶ **주제어:** 양자 컴퓨팅, 노이즈 모델, 양자 게이트 조합, 메타모픽 테스트, 메타모픽 관계

• First Author: Woo-Yeob Hyun, Corresponding Author: Dongsu Kang

*Woo-Yeob Hyun (a16-14112@kaay.ac.kr), Dept. of AI / System Science, Korea Army Academy at Yeongcheon

**Dongsu Kang (dasekang@korea.kr), Dept. of Cyber and Computer Engineering, Korea National Defense University

• Received: 2026. 03. 13, Revised: 2026. 04. 12, Accepted: 2026. 04. 23.

I. Introduction

중첩과 얽힘 같은 양자역학적 특성을 활용한 양자 컴퓨팅은 기존 컴퓨팅의 연산 능력을 뛰어넘을 기술로 주목받고 있다[1]. 그러나 양자 컴퓨팅의 기본 정보 단위인 큐비트는 외부 환경과 노이즈에 민감하여, 동일한 회로를 반복 실행하더라도 출력 확률 분포가 매번 달라질 수 있다. 이로 인해 동일한 출력 분포를 예상할 수 있는 게이트 조합이라도 노이즈 환경에서는 일관된 출력 분포를 보장하기 어려운 문제가 발생한다[2].

양자 테스트 연구들은 Qiskit과 같은 양자 시뮬레이션 플랫폼의 컴파일러 오류를 탐지하거나, 특정 알고리즘의 동작을 단일 노이즈 모델 환경에서 검증하는 데 집중하였다. 이는 기능적 정확성을 확인하는 데 유효하지만, 동일한 출력 분포를 예상하는 게이트 조합이 각기 다른 물리적 오류 특성을 가진 노이즈 환경에서 어떻게 반응하는지에 대한 체계적인 분석을 다루지 못하는 한계가 있다[3].

본 연구는 이상적인 환경에서 동일한 출력 분포를 예상할 수 있는 양자 게이트 조합을 메타모픽 테스트 케이스로 설계하고, 세 가지 노이즈 모델을 적용하여 각 노이즈가 출력 확률 분포에 미치는 영향을 분석한다. 이를 위해 메타모픽 테스트 기법을 도입하여 노이즈 환경에서 출력 간의 논리적 관계인 메타모픽 관계가 유지되는지를 Chi-Square 검정으로 판정하고, KL-Divergence와 Cosine Similarity를 통해 예상 출력 분포와의 차이가 노이즈 강도에 따라 누적되는 과정을 정량적으로 추적한다.

본 논문의 구성은 다음과 같다. 2장에서는 양자 컴퓨팅 및 양자 게이트, 양자 노이즈 모델, 메타모픽 테스트에 대해 소개한다. 3장에서는 메타모픽 테스트에 기반한 테스트 케이스를 설계하고 평가 지표를 제시한다. 4장에서는 설계한 테스트 케이스를 Qiskit Aer 시뮬레이터를 이용해 실험하고, 다양한 노이즈 모델 환경에서의 출력 분포를 평가 지표를 통해 분석한다. 5장에서 결론을 맺는다.

II. Related Works

2.1 Quantum Computing and Gate

양자 컴퓨팅은 큐비트를 정보의 기본 단위로 사용한다. 0 또는 1의 값을 갖는 고전 컴퓨팅의 비트(Bit)와 달리, 큐비트는 아직 결정되지 않은 상태인 중첩과 둘 이상의 큐비트가 서로 강하게 연결되는 얽힘, 측정하는 순간 중첩 상태가 무너지고 0 또는 1의 값이 결정되는 측정의 고유한

양자 역학적 특성을 활용한다.

양자 연산은 양자 게이트를 통해 수행되는데, 이는 큐비트의 상태를 변화시키는 연산 단위이다. 단일 큐비트에 작용하는 게이트에는 큐비트를 중첩 상태로 만드는 Hadamard(H) 게이트, 큐비트 값을 뒤집는 Pauli-X 게이트, 위상을 바꾸는 Pauli-Z 게이트 등이 있다. 다중 큐비트 게이트인 CNOT 게이트는 두 큐비트 간의 얽힘을 생성하는 역할을 한다. 기존 컴퓨팅과 양자 컴퓨팅을 <표 1>과 같이 비교한다.

Table 1. Comparison of Quantum Computing and Classical Computing

Category	Classical Computing	Quantum Computing
Information Unit	Bit	Qubit
Operation Principle	Logic Gate (Sequential Processing)	Quantum Gate (Parallel Processing)
Processing Ability	Limited Parallelism	Massive Parallelism
Calculation Result	Deterministic	Probabilistic
Sensitivity	High Stability	Sensitive to Environment
Realization	Transistor, CPU	Superconducting Qubit

2.2 Quantum Noise Model

큐비트는 외부 환경에 민감하여 노이즈로 인해 상태가 쉽게 손상된다. 이러한 물리적 오류 현상을 시뮬레이션으로 모사하기 위해 다양한 노이즈 모델들이 사용된다[4-5]

Depolarizing 노이즈 모델은 가장 일반적이고 단순한 형태의 모델로, 큐비트의 상태를 일정한 확률로 완전히 무작위적인 상태로 만든다. 2-큐비트가 가질 수 있는 4가지 출력 상태(|00>, |01>, |10>, |11>)의 확률이 점차 균등해지는 경향을 보이며, 노이즈 강도가 커질수록 이 현상은 뚜렷하게 나타난다. 이는 예상 출력 분포가 논리적인 출력값을 잃고 전반적으로 일반화됨을 의미한다.

Amplitude Damping 노이즈 모델은 큐비트가 에너지를 잃는 물리적 현상을 모사한다. 큐비트의 높은 에너지 상태(|1>)가 시간이 지남에 따라 낮은 에너지 상태(|0>)로 수렴하는 특성으로 인해, 2-큐비트 출력 분포 중 |1>을 포함하는 상태의 확률은 점차 감소하고, |00> 상태의 확률은 점차 증가하는 비대칭적 왜곡이 나타난다.

Phase Damping 노이즈 모델은 큐비트의 에너지(진폭)에는 영향을 주지 않으면서 양자 상태 간의 상대적 위상 정보만을 점차 손실시킨다. 양자 계산의 핵심인 중첩과 얽

힘은 위상 관계에 의존하는데, 위상 정보의 손실은 양자 간의 간섭 효과를 사라지게 하여 얽힘 상태를 깨뜨린다. 그 결과 이상적인 예상 출력 상태의 확률이 줄어드는 동시에 나타나지 않아야 할 오류 상태가 나타날 수 있다.

2.3 Metamorphic Testing

일반적인 소프트웨어 테스트는 동일한 입력에 대해 동일한 출력을 기대하는 결정론적 특성을 전제로 한다. 그러나 양자 컴퓨팅은 측정 시점에 확률적으로 결과가 결정되는 비결정적 특성을 가지므로, 동일한 회로를 반복 실행하더라도 출력 확률 분포가 달라질 수 있다. 이로 인해 특정 실행 결과의 정확성을 판별하거나 기대 출력과 비교하는 방식의 테스트 기법을 그대로 적용하기 어렵다. 메타모픽 테스트는 명확한 기대 출력 없이도 입력과 출력 사이의 논리적 관계인 메타모픽 관계(MR, Metamorphic Relation)의 유지 여부를 통해 테스트를 수행[6]할 수 있어, 확률적 출력 분포를 다루는 양자 컴퓨팅 환경에 적합한 테스트 접근 방법이다[7].

MorphQ는 Qiskit 시뮬레이터의 컴파일러 오류를 탐지하는 데 메타모픽 접근을 활용[8]하였으나, 실제 양자 하드웨어에서 발생하는 물리적 노이즈가 출력 분포에 미치는 영향은 포함하지 않았다. Grover 알고리즘의 동작 검증을 위해 입력 변형에 따른 출력 분포 관계를 분석[9]하였으나, 단일 알고리즘 수준에 한정되어 게이트 조합 단위의 체계적 분석으로는 확장되지 않았다.

선행 연구의 제한사항을 보완하여, 이상적 환경에서 논리적으로 동일한 출력 분포를 예상할 수 있는 2-큐비트 게이트 조합을 메타모픽 테스트 케이스로 체계적으로 설계하고, 세 가지 이질적 노이즈 모델을 동시에 적용하여 각 노이즈의 물리적 오류 특성을 Chi-Square, KL-Divergence, Cosine Similarity의 세 가지 정량 지표를 통해 종합적으로 분석한다.

III. Test Case Design

3.1 Defining Metamorphic Relationship

메타모픽 관계는 2-큐비트 게이트 조합을 대상으로 정의하고, 여러 노이즈 모델에서 관계 유지 여부를 정량적으로 평가한다[10]. 본 연구에서는 ‘논리적으로 동일한 출력 분포를 예상할 수 있는 게이트 조합은 노이즈가 없는 이상적인 환경에서 동일한 출력 분포를 가지며, 노이즈 환경에서도 이 관계는 유지되어야 한다’로 메타모픽 관계를 정의

한다. 정의한 관계가 다양한 노이즈 환경에서 얼마나 잘 유지되는지, 그리고 각 노이즈가 예상 출력 분포를 어떻게 변화시키는지를 실험을 통해 정량적으로 측정한다. 이를 위해 정의한 메타모픽 관계를 기반으로 한 6개의 테스트 케이스를 <표 2>와 같이 설계하였다.

Table 2. Test Cases based on Metamorphic Relation

Case	Gate Combination Configuration	Expected Output
TC1	Base Combination (H + CNOT) H(q[0])→CNOT(q[0],q[1])	00> 50% 11> 50%
TC2	Logical Equivalence Add X gate to q[0] X(q[0])→H(q[0])→CNOT(q[0],q[1])	
TC3	Single Phase Flip Add Z gate to q[0] Z(q[0])→H(q[0])→CNOT(q[0],q[1])	
TC4	Double Phase Flip Add Z gate to q[0], q[1] Z(q[0],q[1])→H(q[0])→CNOT(q[0],q[1])	
TC5	Gate Invalidation Add H gate to q[0] twice H(q[0])→CNOT(q[0],q[1])→2H(q[0])	
TC6	Swap Operations Change order of base gates H(q[1])→CNOT(q[1],q[0])	

소스 게이트 조합(TC1)은 테스트의 기준점으로, 2-큐비트 얽힘 상태를 생성하는 기본적인 조합이다. 첫 번째 큐비트에 Hadamard 게이트를 적용하여 중첩 상태를 만들고, 이후 두 큐비트에 CNOT 게이트로 얽힘 상태를 만든다. 이 현상을 양자 상태 전개로 나타내면 초기 상태 |00>이 얽힘 상태 $\Phi^+ = (|00\rangle + |11\rangle) / \sqrt{2}$ 로 변환되며, 이상적인 환경에서 이 회로는 |00>과 |11>상태를 각각 50% 확률로 출력한다. 팔로우업 게이트 조합(TC2~TC6)은 소스 게이트 조합에 논리적으로 동일한 출력 분포를 유도하는 다양한 구조적 변형을 가한 것으로, 각 변형은 양자역학적 원리에 따라 이상적인 환경에서는 출력 분포에 영향을 주지 않는다. 게이트 추가, 중복, 순서 변경 등의 변형으로 생성하였으며, 소스와 팔로우업을 포함한 6개의 조합 모두 동일한 예상 출력 분포를 기준으로 메타모픽 관계 유지 여부를 판정한다.

각 테스트 케이스에 대한 설명은 다음과 같다. TC2~TC4는 논리 등가 및 위상 반전 관계로, Pauli X, Z 게이트와 Hadamard 게이트의 상호작용을 이용한 변형이다. 전체적인 논리는 유지하면서 특정 큐비트의 위상만을 변경하므로 측정 확률에는 영향을 미치지 않는다. TC2는 H 게이트 이전에 X 게이트를 추가하면 위상 부호만 달라지며, 위상 부호의 차이는 측정 확률에 영향을 주지 않

므로 출력 분포는 TC1과 동일하게 유지된다. TC3는 H 게이트 이전에 Z 게이트를 추가하고, 초기 상태와 변화가 없어 이후 연산 결과가 TC1과 동일하다. TC4는 두 큐비트에 모두 Z 게이트를 적용하더라도 초기 상태가 유지되어 동일한 연산 결과를 예측할 수 있다.

TC5는 Hadamard 게이트를 두 번 연속 적용하여 항등 연산으로 만드는 변형으로, 연산이 논리적으로 무효화되는 관계를 테스트한다. Hadamard 게이트의 연속 적용으로 자기 역원 특성의 효과를 지니고, CNOT 이후 추가된 두 번의 H 게이트는 양자 상태에 아무런 영향을 미치지 않아 최종 출력 분포는 TC1과 동일하게 유지된다. TC6은 CNOT 게이트의 제어 큐비트와 목표 큐비트를 반대로 적용하는 대칭적 변형으로, 얽힘 상태의 대칭성을 활용하여 연산 순서 변경이 출력 분포에 미치는 영향을 평가한다. 얽힘 상태 $|\Phi^+\rangle = (|00\rangle + |11\rangle) / \sqrt{2}$ 는 큐비트의 순서 교환에 대해 불변하는 대칭성을 가지므로, 제어·목표 큐비트의 역할을 바꾸어도 동일한 상태가 생성되어 출력 분포는 TC1과 동일하게 유지된다.

3.2 Evaluation Metrics

설계한 테스트 케이스들을 평가하기 위해 세 가지 평가 지표를 제시한다. 각 노이즈 모델은 시스템의 안정성 저하, 에너지 손실, 위상 손실과 같은 고유한 물리적 오류를 모사하므로, 단일 지표만으로는 이러한 복합적인 영향을 평가하기 어렵다. 본 연구에서는 메타모픽 관계 유지 여부를 판정하는 핵심 지표인 Chi-Square와 예상 출력 분포와의 차이가 노이즈 강도에 따라 누적되는 과정을 추적하기 위한 KL-Divergence와 Cosine Similarity를 보조 지표로 사용한다. 세 지표는 각각 메타모픽 관계의 판정, 분포 이탈의 크기 추적, 분포 구조의 방향성 추적이라는 상호 보완적인 역할을 수행하도록 설계하였다.

먼저 Chi-Square(χ^2) 검정은 관찰 분포(O_i , 테스트 케이스 별 출력되는 결과)와 기대 분포(E_i , $|00\rangle$ 과 $|11\rangle$ 각 50% 출력) 간 차이를 기반으로 두 분포의 통계적 동일 여부를 판단하여, 메타모픽 관계의 유지 여부를 판정한다. 통계량(χ^2)은 두 분포 간의 불일치 정도를 나타내는 단일 값으로 식 (1)과 같이 계산된다[11].

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Chi-Square 검정을 핵심 지표로 선택한 이유는 양자 회로의 출력이 이산 확률 분포 형태로 나타나며, 제한된 측정 횟수 내에서 관찰된 빈도와 기대 빈도 간의 통계적

유의성을 검정하는 데 적합하기 때문이다. 또한, p-value를 통해 메타모픽 관계 유지 여부를 명확한 기준으로 판정할 수 있어, 노이즈 환경에서의 관계 붕괴 시점을 객관적으로 결정할 수 있다.

값을 해석하기 위해 p-value를 사용하며, p-value의 통계적 유의수준(α)을 0.05로 설정한다. p-value 값에 따라 메타모픽 관계 유지 상태를 세분화한 기준은 <표 3>과 같다.

Table 3. Criteria for Determining based on p-value

p-value	Description	Criteria
$p > 0.5$	Highly Similar	Maintain MR
$0.05 < p \leq 0.5$	Uncertain Similarity	Uncertain MR
$p \leq 0.05$	Significant Difference	MR Collapse

p-value는 두 분포가 실제로는 동일하다고 가정할 때, 측정된 불일치 수준(χ^2) 이상이 순전히 우연에 의해 관찰될 확률을 의미하며, χ^2 분포를 통해 계산된다. 따라서 p-value가 낮을수록 관찰된 차이가 우연의 결과일 가능성이 희박함을 의미하며, $p \leq 0.05$ 이면 실제 출력 분포가 예상 출력 분포에서 통계적으로 유의미하게 이탈한 것으로 판단하여 메타모픽 관계가 붕괴된 것으로 판정한다.

KL(Kullback-Leibler) Divergence는 이상적인 예상 출력 분포 P를 기준으로, 노이즈 적용 후 실제 출력 분포 Q가 예상 분포에서 얼마나 이탈하였는지를 정보량 차이로 정량화하며[12], 두 분포간 정보량 차이는 식 (2)와 같이 정의된다[13].

$$KL(P \parallel Q) = \sum_x P(x) \cdot \log\left(\frac{P(x)}{Q(x)}\right) \quad (2)$$

KL-Divergence를 보조 지표로 선택한 이유는 Chi-Square 검정이 메타모픽 관계의 붕괴 여부를 이진으로 판정하는 데 그치는 반면, KL-Divergence는 붕괴 이후에도 노이즈 강도가 증가함에 따라 예상 분포와의 차이가 얼마나 누적되는지를 연속적으로 추적할 수 있기 때문이다. 값이 0에 가까울수록 실제 출력이 예상 분포에 근접하고, 값이 클수록 예상 분포에서의 이탈이 크다.

Cosine Similarity는 예상 출력 분포와 실제 출력 분포를 각각 벡터로 간주하고, 두 벡터 간의 방향성(각도)을 이용해 구조적 패턴의 유사도를 측정한다[14]. 두 분포 벡터 간의 구조적 패턴 유사도는 식 (3)과 같이 측정된다[15].

$$CS = \cos(\theta) = \frac{P \cdot Q}{\|P\| \cdot \|Q\|} \quad (3)$$

Cosine Similarity를 보조 지표로 선택한 이유는

KL-Divergence가 분포 이탈의 크기를 추적하는 반면, Cosine Similarity는 분포의 전체적인 구조가 어느 방향으로 얼마나 변형되었는지를 직관적으로 파악할 수 있어 두 지표가 상호 보완적인 역할을 수행하기 때문이다. 값이 1에 가까울수록 실제 출력 분포의 구조가 예상 분포와 유사하고, 0에 가까울수록 분포의 구조적 왜곡이 심화되었음을 나타낸다.

세 지표를 함께 활용함으로써 Chi-Square로 메타모픽 관계의 붕괴 시점을 판정하고, KL-Divergence와 Cosine Similarity로 붕괴 이후 노이즈로 인한 분포 왜곡의 크기와 방향을 동시에 추적할 수 있어 노이즈 모델별 영향을 종합적으로 평가할 수 있다.

IV. Experiment and Analysis

4.1 Experiment Environment

IBM의 양자 컴퓨팅 프레임워크인 Qiskit Aer를 활용하여 실험하였다. 각 테스트 케이스에 세 가지 노이즈 모델을 각각 적용하였으며, 노이즈 강도는 0.0(오류 없음)부터 0.1까지 0.01 간격으로 점진적으로 증가시켜 노이즈 강도에 따른 출력 분포의 변화와 메타모픽 관계의 변화를 세밀하게 관찰할 수 있도록 설계한다.

특히, 트랜스파일 최적화 레벨을 0으로 설정하여 Qiskit 시뮬레이터가 내부적으로 게이트 조합을 최적화(단순화)하는 것을 방지하고, 설계된 구조적 차이가 결과에 그대로 반영되도록 통제하였다. 모든 실험은 동일한 환경에서 수행되었으며, 주요 실험 환경은 <표 4>와 같다.

6개의 테스트 케이스를 세 가지 노이즈 모델 환경에서 실행한 결과를 출력 확률 분포와 세 가지 평가 지표로 시각화하여 분석하였다. 각 노이즈의 물리적 특성을 명확히 관찰할 수 있도록 측정 기저를 달리 설정하였다.

Depolarizing과 Amplitude Damping 노이즈 모델은 에너지 준위 변화를 측정하기 위해 표준 측정 방식인 Z-기저(Z-Basis)에서, Phase Damping 노이즈 모델은 위상 정보 손실을 명확히 확인하기 위해 X-기저(X-Basis)에서 측정하였다. Z-기저는 큐비트의 에너지 상태($|0\rangle$, $|1\rangle$)를 구분하는 표준 측정 방식으로, 에너지 준위 변화를 수반하는 두 노이즈 모델의 영향을 가장 직접적으로 관찰할 수 있다. 반면 Phase Damping 노이즈는 에너지 변화 없이 위상 정보만을 손실시키므로, 위상 차이에 민감한 X-기저 측정을 통해 간섭 효과의 감쇠를 보다 명확하게 확인할 수 있다.

Table 4. Experimental Environment

Category	Specification
Platform	Google Colab
Simulator	Qiskit Aer QASM Simulator
Noise Model	Depolarizing, Amplitude Damping, Phase Damping
Noise Level	0.0~0.1(in increments of 0.01)
Shots	4,096 times for each noise level
Simulator Seed	1,234 (to ensure consistency of results)

4.2 Results and Analysis

<Fig 1>은 노이즈 강도에 따른 각 테스트 케이스의 출력 확률 분포 변화를 나타낸다. 실험 결과 각 노이즈 모델은 고유한 물리적 오류 특성에 따라 서로 다른 방식으로 출력 확률 분포를 예상 출력 분포에서 이탈시켰다. 모든 실험에서 공통적으로 예상 출력 확률 ($|00\rangle$ 과 $|11\rangle$ 각 50%)에서 벗어나는 편차의 크기가 세 가지 평가 지표의 변화와 직접적으로 연결되었으며, 출력 분포가 50% 균형점에서 멀어질수록 메타모픽 관계의 변화 정도를 나타내는 지표 값도 크게 변화했다.

Depolarizing 노이즈 모델(A1~A6)은 $|00\rangle$ 과 $|11\rangle$ 의 확률이 50%에서 서서히 균등 분포 방향으로 낮아지며, 나타나지 않아야 할 $|01\rangle$ 과 $|10\rangle$ 상태의 확률이 점차 증가하는 양상이 관찰되었다. 이는 노이즈 강도가 증가할수록 얽힘 구조가 균일하게 소실되어 4개의 출력 상태가 각 25%로 수렴하는 과정으로 해석된다. Amplitude Damping 노이즈 모델(B1~B6)은 $|11\rangle$ 확률이 줄어들고 $|00\rangle$ 확률이 증가하여 50% 균형점에서 가장 빠르고 비대칭적으로 멀어졌다. 두 큐비트 모두 에너지 이완의 영향을 이중으로 받으므로, 에너지 구조의 직접적 파괴로 인한 분포 왜곡이 세 가지 노이즈 모델 중 가장 집중적으로 발생한다. Phase Damping(C1~C6) 노이즈 모델은 X-기저 측정 결과, 에너지 손실 없이 위상 정보만을 점차 손실시키는 결어긋남 현상으로 인해 $|00\rangle$ 과 $|11\rangle$ 확률이 50%에서 점진적으로 멀어지는 모습을 보였으며, 두 노이즈 모델과 비교하였을 때 중간 수준의 편차율을 나타냈다.

<Fig 2>는 세 가지 평가 지표를 통해 노이즈 강도에 따른 메타모픽 관계의 변화를 나타낸다. 모든 노이즈 모델에서 p-value는 노이즈 강도가 약 0.01에 도달하는 시점을 전후하여 유의수준 이하로 하락하며 메타모픽 관계의 붕괴를 보였다. 이는 소량의 노이즈만으로도 실제 출력 분포가 예상 출력 분포에서 통계적으로 유의미하게 이탈할 수 있음을 의미하며, 현실의 양자 하드웨어 환경에서 노이즈 제어의 임계값 설정이 필요함을 나타낸다.

메타모픽 관계 붕괴 시점은 세 가지 노이즈 모델에서 유사하게 나타났으나, 붕괴에 이르는 과정은 각 노이즈가 출력 분포를 변형시키는 매커니즘에 따라 차이를 보였다. Depolarizing 노이즈 모델은 KL-Divergence의 선형적 증가와 Cosine Similarity의 완만한 감소로 나타나 분포 왜곡이 전반적으로 균등하게 누적됨을 보여주었다. Amplitude Damping 노이즈 모델은 KL-Divergence가 가장 가파르게 증가하고 Cosine Similarity가 가장 큰 폭으로 하락하여, 예상 출력 분포와의 차이가 세 가지 노이즈 모델 중 가장 빠르고 크게 누적되었다. Phase Damping 노이즈 모델은 두 노이즈 모델의 중간 수준으로 지표가 변화하였다.

설계한 6개의 테스트 케이스는 구조적으로 서로 다름에도 불구하고, 동일한 노이즈 모델 내에서 거의 동일한 출력 분포 변화 패턴과 지표 변화를 보였다. 이는 게이트 조합의 미세한 구조적 차이보다 시스템에 가해진 노이즈의 종류와 물리적 특성이 출력 분포 변화와 메타모픽 관계 붕괴를 결정하는 지배적 요인임을 나타내며, 메타모픽 테스트 기법이 노이즈 유형별 시스템 신뢰성 평가 도구로서 유효함을 보여준다.

V. Conclusions

본 연구는 이상적인 환경에서 동일한 출력 분포를 예상할 수 있는 2-큐비트 양자 게이트 조합을 메타모픽 테스트 케이스로 설계하고, Depolarizing, Amplitude Damping, Phase Damping 노이즈 모델을 적용하여 각 노이즈가 출력 확률 분포에 미치는 영향을 분석하였다. 이를 위해 메타모픽 테스트 기법을 적용하여 Chi-Square, KL-Divergence, Cosine Similarity 세 가지 지표를 종합적으로 활용하였다.

실험 결과, 모든 노이즈 모델에서 노이즈 확률이 0.01에 도달하는 시점에 p-value가 유의수준 이하로 하락하며 메타모픽 관계가 붕괴됨을 확인하였다. 이는 소량의 노이즈 만으로도 동일한 출력 분포를 예상했던 게이트 조합들의 실제 출력이 예상 분포에서 통계적으로 유의미하게 이탈할 수 있음을 보여준다. 또한 KL-Divergence와 Cosine Similarity를 통해 메타모픽 관계 붕괴 이후에도 노이즈 강도가 증가함에 따라 예상 분포와의 차이가 지속적으로 누적됨을 정량적으로 확인할 수 있었다.

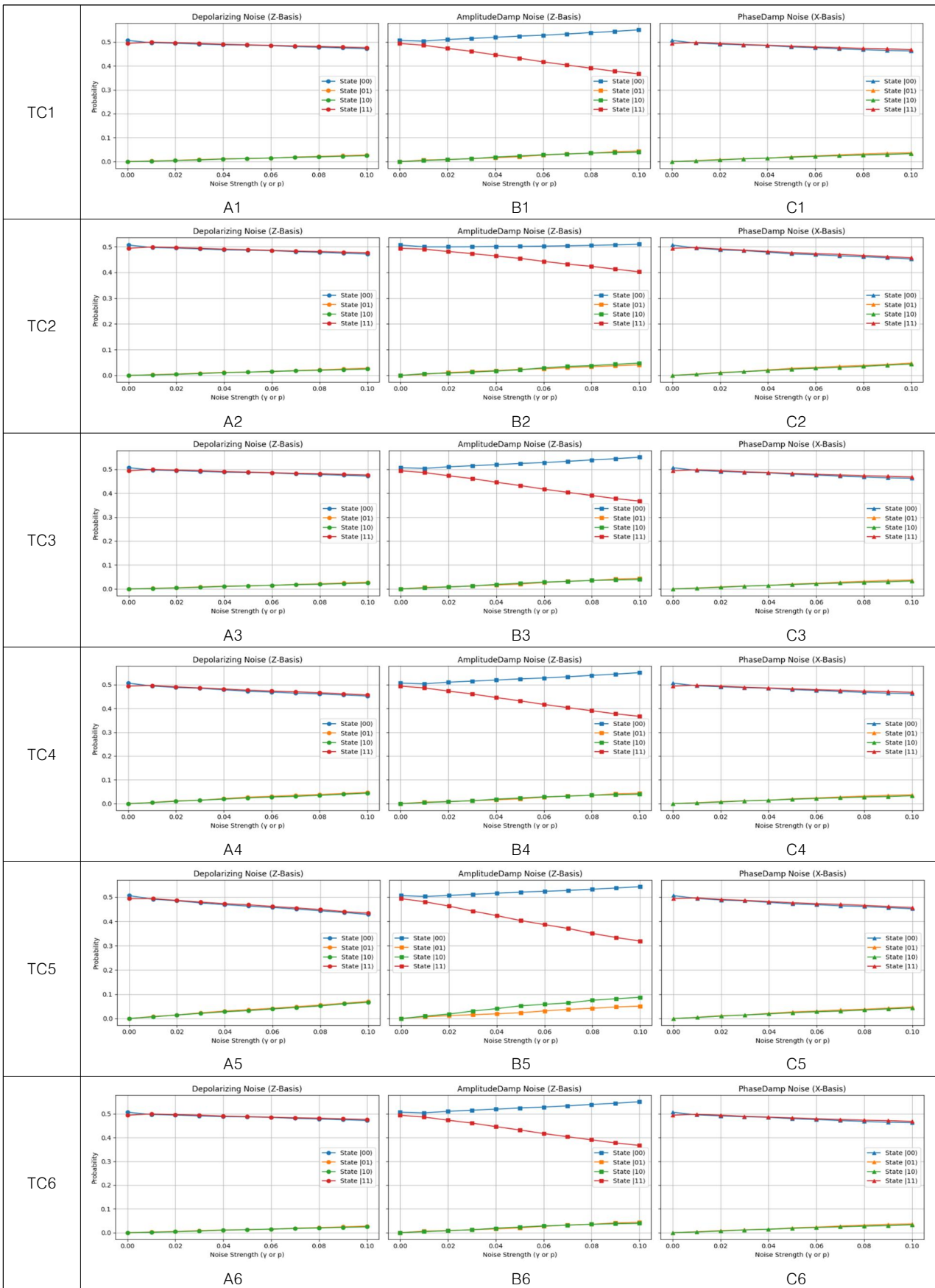


Fig. 1. Change in Test Case Output Distribution by Noise Model

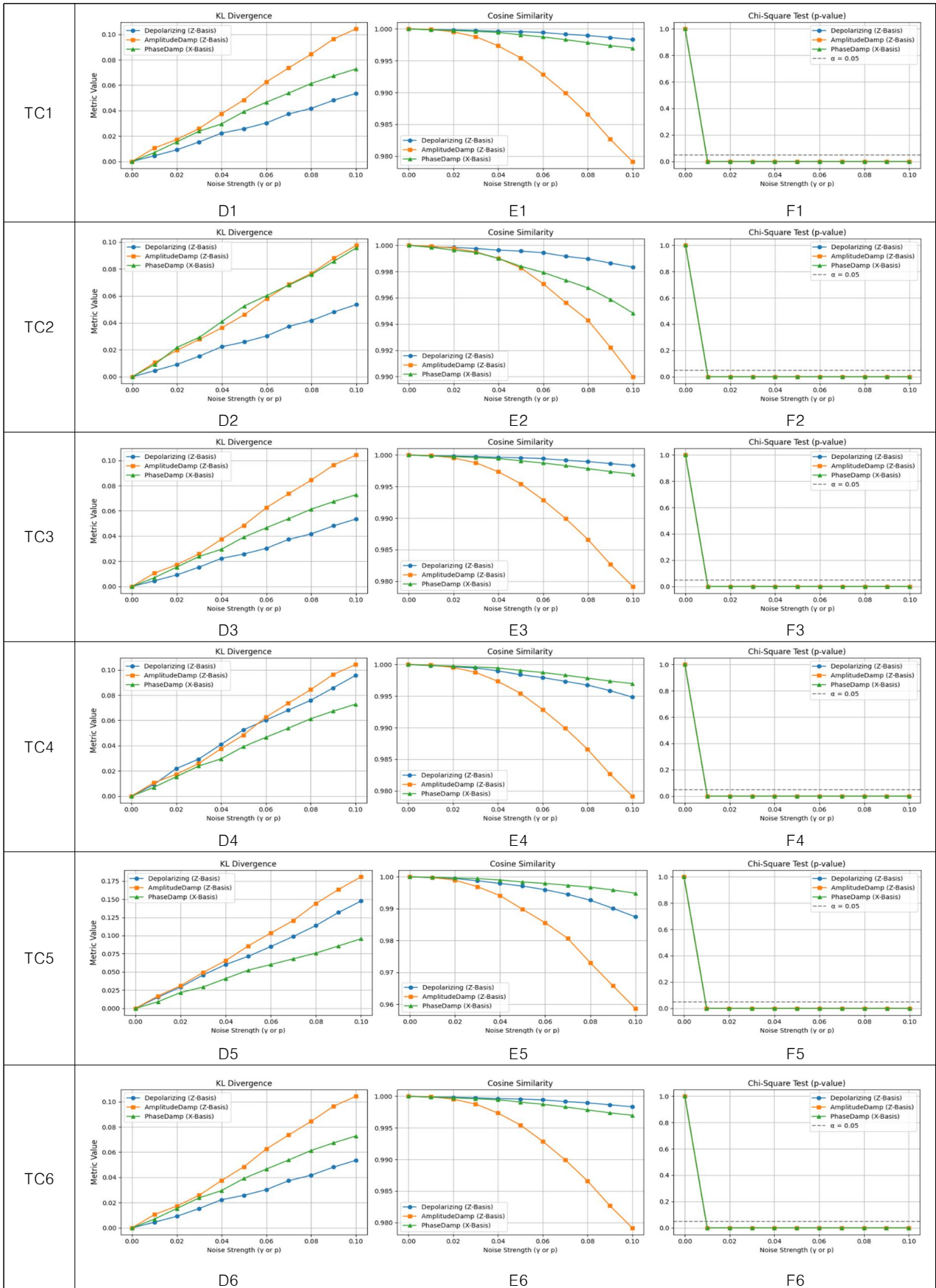


Fig. 2. Metamorphic Relation using Evaluation Metrics

ACKNOWLEDGEMENT

This paper is based on the first author's master's thesis entitled 'Design and Evaluation of Metamorphic Test Cases using Quantum Gate Combinations' (2026) at Korea National Defense University.

And This paper was supported by the Chungseongdae Research Institute of Korea Army Academy at Yeongcheon(KAAY) through the 2026 Paper Publication Grant.

REFERENCES

- [1] W. Hyun, "Design and Evaluation of Metamorphic Test Cases using Quantum Gate Combinations," M.S. thesis, Dept. of Computer Engineering, Korea National Defense University, Nonsan, Korea, Jan. 2026. (in Korean)
- [2] S. Resch and U. R. Karpuzcu, "Benchmarking Quantum Computers and the Impact of Quantum Noise," *ACM Computing Surveys*, Vol. 54, No. 7, pp. 1-35, Jul. 2021.
- [3] M. L. O. Atencio, L. Lamata, and J. C. Pascual, "Impact of Amplitude and Phase Damping Noise on Quantum Reinforcement Learning: Challenges and Opportunities," *The European Physical Journal Special Topics*, Vol. 234, No. 7, pp. 1-11, Jul. 2025.
- [4] A. S. Roy, C. V. Namitha, S. Mukerjee, P. K. Panigrahi, and U. Sinha, "Intraparticle Entanglement in Noisy Quantum Channels: Degradation and Revival through Amplitude Damping," *Frontiers in Quantum Science and Technology*, Vol. 4, pp. 1-12, Jun. 2025.
- [5] T. Ahmed, A. Marchisio, M. Kashif, and M. Shafique, "Noisy HQNNs: A Comprehensive Analysis of Noise Robustness in Hybrid Quantum Neural Networks," *Proceedings of the International Joint Conference on Neural Networks*, pp. 1-10, Jul. 2025.
- [6] D. Kang, "Bridging Fuzz Testing and Metamorphic Testing for Classification of Machine Learning," *Proceedings of the 2022 IEEE International Conference on Consumer Electronics*, pp. 1-2, Jan. 2022.
- [7] I. Lee and D. Kang, "Generating Metamorphic Test Cases for Transformer-based Korean Summary," *KIISE Transactions on Computing Practices*, Vol. 29, No. 11, pp. 509-517, Nov. 2023. (in Korean)
- [8] M. Paltenghi and M. Pradel, "MorphQ: Metamorphic Testing of the Qiskit Quantum Computing Platform," *Proceedings of the 45th IEEE/ACM International Conference on Software Engineering*, pp. 1-12, May. 2023.
- [9] R. Abreu, J. P. Fernandes, L. Llana, and G. Tavares, "Metamorphic Testing of Oracle Quantum Programs," *Proceedings of the 3rd International Workshop on Quantum Software Engineering*, pp. 16-23, May. 2022.
- [10] Y. Kim and D. S. Kang, "Generation Composite Variable Rules for Metamorphic Testing of CNN Image Classification Models," *KIISE Transactions on Computing Practices*, Vol. 30, No. 2, pp. 77-83, Feb. 2024. (in Korean)
- [11] Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," *Proceedings of the 14th International Conference on Machine Learning*, pp. 412-420, Jul. 1997.
- [12] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 79-86, Mar. 1951.
- [13] J. R. Hershey and P. A. Olsen, "Approximating the Kullback-Leibler Divergence Between Gaussian Mixture Models," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 317-320, Apr. 2007.
- [14] K. You, "Semantics at an Angle: When Cosine Similarity Works Until It Doesn't," *arXiv preprint arXiv:2504.16318*, Apr. 2025.
- [15] N. Dehak, R. Dehak, J. Glass, D. Reynolds, and P. Kenny, "Cosine Similarity Scoring without Score Normalization Techniques," *Proceedings of the Speaker and Language Recognition Workshop*, pp. 1-5, Jun. 2010.

Authors



Woo-Yeob Hyun received the B.S. degree in Electronic Engineering from Daejin University in 2012 and the M.S. degree in Computer Engineering from the National Defense University in 2026.

Prof. Hyun joined the faculty of the Department of AI and System Science at Korea Army Academy at Yeongcheon, Yeongcheon, Korea, in 2026. He is interested in SW testing, quantum computing, cyber warfare, and AI systems.



Dongsu Kang received his B.S. in Electrical Engineering from the Republic of Korea Naval Academy and his Ph.D. in Computer Science and Engineering from Korea University.

He is professor in Department of Cyber and Computer Engineering, Korea National Defense University (KNDU). His primary research areas include analysis of North Korea's ICT infrastructure, with a focus on the Red Star OS and science and technology trends, using data science and software security testing methodologies, as well as defense AI and software vulnerability analysis. He is ranked in the Top 1% of Computer Science researchers on DBpia, South Korea's leading academic database platform.