China Instrument and Control Society
Instrumentation
OPEN ACCESS
*INSTRUMENTATION*, Vol. 12, No. 1, March 2025 https://doi.org/10.15878/j.instr.202500223

Article

# An Improved Copula-Based Test Selection Design Strategy for Fault Detection and Isolation Based on PSO

**Xiuli Wang[1,2], Dongdong Xie[1], Yang Li[3*], Chun Liu[3], Xinyu Hu[4]**

[1] College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China

[2] Zhejiang Hengjiu Transmission Technology INC.,LTD., Zhuji, Zhejiang 311800, China

[3] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China

[4] State Grid Nantong Power Supply Company, Nantong, Jiangsu 210000, China

[*] Corresponding author email: yongerli@shu.edu.cn

**Abstract:** Test selection design (TSD) is an important technique for improving product maintainability, reliability and reducing lifecycle costs. In recent years, although some researchers have addressed the design problem of test selection, the correlation between test outcomes has not been sufficiently considered in test metrics modeling. This study proposes a new approach that combines copula and D-Vine copula to address the correlation issue in TSD. First, the copula is utilized to model FIR on the joint distribution. Furthermore, the D-Vine copula is applied to model the FDR and FAR. Then, a particle swarm optimization is employed to select the optimal testing scheme. Finally, the efficacy of the proposed method is validated through experimentation on a negative feedback circuit.

**Keywords:** Design of testability; fault detection and isolation (FDI); copula function; vine copula model; particle swarm optimization (PSO)



**Citation:** Xiuli Wang, Dongdong Xie, Yang Li, Chun Liu, Xinyu Hu."An Improved Copula-Based Test Selection Design Strategy for Fault Detection and Isolation Based on Pso." Instrumentation 12, no.1 (March 2025). https://doi.org/10.15878/j.instr.202500223

## 1 Introduction

The rapid advancement of technology and increased system integration have made high-tech equipment more complex. This complexity leads to fault detection and isolation (FDI) even if predictive maintenance is more challenging and costly[1,2]. As is known, efficient FDI of complex systems requires comprehensive, accurate, and effective information perception as a prerequisite[3-5]. The test selection design (TSD) plays a crucial role in providing fault information and helps reduce life cycle costs[6,7]. Since the 1980s, TSD has been vital for determining system states and isolating faults by the U.S. military and research institutions. Although China began focusing on TSD technology later, significant strides have been made since the 1995 Outline of Equipment Testability, leading to increased attention from Chinese universities and research organizations[8,9].

TSD modeling is critical in analyzing testability and assessing test performance[10]. Based on TSD criteria, researchers have recently developed a range of modeling methods, which include mathematics-based approaches, electronic design automation (EDA) -base techniques, discrete-event-triggered methods, and information correlation-based modeling strategies[11-14]. Mathematics-based test modeling methods typically involve constructing mathematical models correlating tests with faults[11]. While it can deliver reliable analytical results, particularly in false alarm analysis, it may only encompass a limited range of known fault modes. Each fault mode requires the formulation of a unique mathematical model, which poses significant challenges for system information integration. The EDA-based modeling approach employs EDA techniques and tools to automatically generate test vectors compatible with automated test equipment[12].

Additionally, it utilizes these resources to automatically diagnose the root causes of faults resulting in component failure. However, this method demands a thorough understanding of circuit design. The discrete event-triggered modeling approach describes TSD system operations through event-driven sequences occurring over discrete time intervals. However, this method faces limitations in analyzing the logic based on values between fault signals and tests. Consequently, it cannot simulate test logic and assess test performance effectively[13].

In recent years, information correlation-based modeling methods have garnered increasing attention due to their ability to intuitively reflect the system's operational status at each test point. These methods include correlation modeling, multi-signal flow graph modeling, multi-state flow modeling, and information flow modeling. The core principle of these methods involves constructing a fault-to-test correlation matrix, which serves as the foundation for designing testability strategies. Typically, the fault correlation matrix is developed assuming a specific fault state, implying deterministic test outcomes (perfect testing) [15]. However, actual application systems often experience uncertainty in test outcomes due to diverse external and internal factors like electromagnetic interference, measurement uncertainty, and operational variability. Building upon this premise, Zhang et al.[16] introduced a TSD modeling approach grounded in statistical analysis, specifically addressing test result uncertainty using the Bernoulli distribution. Usually, TSD leveraging the Bernoulli distribution operates under the assumption that test outcomes constitute independent random variables. However, actual systems often feature interconnected and interdependent components, leading to correlations among test variables. This renders the assumption of independence among test variables inadequate. Leveraging the statistical model of testability, many researchers delve into various related challenges, including the testability growth problem, testability problem, and testability evaluation problem[17-19].

Recognizing the importance of accounting for test correlation, Ye et al.[15] proposed a method incorporating copulas to model the joint distribution of test results, thereby addressing the correlation among test variables. Expanding on this foundation, Li et al. extend their considerations to encompass the impact of random errors and multiple failure modes by a testability model grounded in joint distribution. Tang et al. [22] introduced a novel perspective by modeling testability using the kernel density estimation (KDE) approach. They then delved deeper into this method, exploring a data-driven TSD modeling approach derived from KDE[23]. On a similar note, Li et al.[24] put forth a fresh data-driven TSD-based modeling approach, incorporating the Kullback-Leibler (KL) divergence.

In test correlation studies, the methodologies described above often rely on copula-based functions. However, employing a multivariate copula function solely to compute the correlation of random variables overlooks nuances such as variations in tail correlation among different pairs of test variables. To solve the multidimensional correlation problems in engineering design, researchers[25,26] proposed the idea of constructing pair-copula construction (PCC) and successively developed the regular vine copula model theory. Jiang et al. [27] proposed a regular vine copula-based structural reliability analysis method to quantify the correlations for complex multidimensional engineering problems. Li et al. [28] proposed a one-step Bayesian copula model selection-assisted D-vine sampling method (OBCS-D) to achieve the coupling of uncertain variables in composites. Consequently, D-Vine copula theory can be used to tackle the correlating test issue. It is important to note that the D-Vine copula theory is founded on continuous data[29]. When constructing fault isolation rate (FIR) models based on the joint distribution, specific transformations are often required for the test outcomes and thresholds[15]. These transformations can potentially disrupt the continuity of the data, which may consequently impact the performance of FIR models under D-Vine copula theory.

Motivated by the above discussions, this paper introduces an innovative approach that combines copula and D-Vine copula to address the correlation issue in testability design. The correlation issue of TSD is leveraging copula to calculate the FIR and D-Vine copula and determine the fault detection rate (FDR) and the false alarm rate (FAR). It is noteworthy that the FAR can be similarly described as the fault detection rate of the system in a fault-free state[15]. Then, the corresponding constraints are modeled based on FDR and FIR. The final optimal test point is then determined using the particle swarm optimization (PSO). The contributions of this paper are described below:

1. A copula function is employed in FIR to effectively characterize the correlation between test outcomes.
2. A new D-Vine copula-based TSD modeling methodology is proposed for modeling the FDR constraint metric to better address the dependence between test outcomes.
3. A PSO algorithm obtains the optimal test set based on the proposed TSD modeling method.

The main organization of this paper is as follows. In Section II, a framework related to TSD modeling is presented. In Section III, the FIR and FDR model are constructed based on the copula theory and vine copula theory, respectively. The corresponding test selection model is built and the PSO algorithm is used to select the optimal test set in Section IV. In Section V, a negative feedback circuit (NFC) is utilized to verify the validity of the proposed method. Finally, Section VI gives the corresponding conclusions.

## 2 TSD framework

Some definitions are given as bellow:

- The set of possible faults is given by $F=\{f_0,f_1,\cdots f_m\}$, for a system, where $m$ denotes the number of faults; and $f_0$ represents the fault-free state.
- The prior probability of the corresponding fault in the fault set $F$ is $P=\{p_0,p_1,\cdots p_m\}$, where $\sum_{i=0}^{m} p_i=1$.
- The set of all tests is depicted by $T=\{t_1, t_2, \cdots, t_n\}$ and the corresponding test cost is presented by $CT=\{ct_1, ct_2, \cdots, ct_n\}$.
- The set of test thresholds is $thr=(thr_1, thr_2, \cdots, thr_n)$..
- The test selection vector is $S=(s_1, s_2, \cdots, s_n)$. If test $t_j$ is selected then $s_j=1$, otherwise $s_j=0$.

Typically, two metrics, FDR (encompassing the system fault detection rate in a fault-free state, i.e., FAR) and FIR serve as indicators of the performance of a system test. The FDR and FIR are calculated based on the test outcomes, thus the outcomes of the tests directly influence both metrics. With the escalating complexity of systems, the components within the system are not isolated entities but intricately interconnected. This interdependence is also reflected in the system's test outcomes, often showcasing correlations. The issue of correlation between test outcomes under fault conditions is illustrated by an example of a linear voltage divider in Fig.1[24].

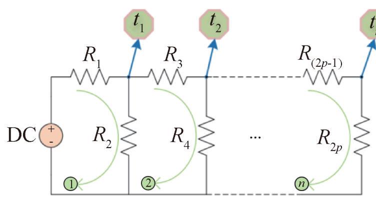


Fig.1 Linear voltage divider

The circuit shown in Fig. 1 consists of a $2p$ potential fault and $n$ test points $\{t_1, t_2, \cdots, t_n\}$. When analyzing the $t_1$ and $t_2$, the voltage values for $t_1$ and $t_2$ can be calculated as Eq.(1) and Eq.(2).

$$V(t_1)=\frac{R_1(R_3+R_4)}{R_1R_2+(R_1+R_2)(R_3+R_4)}V_{dc} \tag{1}$$

$$V(t_2)=\frac{R_1R_3}{R_1R_2+(R_1+R_2)(R_3+R_4)}V_{dc} \tag{2}$$

where $V_{dc}$ represents the voltage source. If a fault is considered according to Eq. (1) and Eq. (2), it can be observed that when the value of $R_1$ changes until the test outcomes for $V(t_1)$ and $V(t_2)$ exceed their respective thresholds, both test points $t_1$ and $t_2$ are capable of detecting $R_1$ faults. This redundancy in testing leads to increasing test costs and re-source wastage. Thus, selecting appropriate test points to mitigate correlation issues for testing becomes essential.

However, relying solely on the multivariate copula function to assess the correlation of random variables has clear limitations[15], particularly regarding neglecting variations in tail correlation across different variable combinations. Here, the D-Vine copula theory is a viable solution based on continuous data. Usually, specific transformations become necessary for test outcomes and thresholds, when establishing FIR models based on joint distribution[15]. These transformations could potentially disrupt data continuity, affecting the efficacy of FIR models based on D-Vine copula theory. This paper proposes a novel test selection design approach that integrates copula and D-Vine copula methodologies. The aim is to compute FIR using copula while mitigating correlation issues in the FDR mode through D-Vine copula. As illustrated in Figure 2, the first step involves collecting test outcomes for different fault states at each test point, which will be used as the total sample set. Next, the FIR based on the joint distribution and the FDR based on the D-Vine copula are computed, corresponding to the left and right parts of the figure, respectively. Finally, after constructing the FIR and FDR models, the fitness function (test cost) is determined and the optimal test set is selected using the PSO algorithm.

## 3 TSD model

In this section, copulas are used to model FIR on the joint distribution, and D-Vine copula is utilized to model FDR.

### *3.1 Copula-based FIR Model*

#### 3.1.1 Copula theory for joint distributions

Copula theory, excelling in characterizing correlations among multidimensional random variables. It has been integrated into test selection design to address the challenge of correlation between test results effectively. For two-dimensional random variables $X$ and $Y$, let the marginal distribution functions be $u=F(x)$ and $v=G(y)$ and the corresponding marginal probability density functions (PDF) be $f(x)$ and $g(y)$. The joint distribution function of $X$ and $Y$ is $F(x,y)$, which can be expressed in terms of a two-dimensional copula function $C(u,v;\theta)$ defined on the space $[0,1]^2$, as follows:

$$F(x,y)=C\left(F(x),G(y)\right) \tag{3}$$

$$f(x,y)=\frac{\partial C(F(x),G(y))}{\partial x\partial y}=c(F(x),G(y))f(x)g(y) \tag{4}$$

The two-dimensional case can be easily generalized into a multidimensional case. Let the n-dimensional random variable be $X=(X_1,X_2,\cdots,X_n)$. The corresponding marginal distribution function is presented by $F_i(x_i)$. $c(\cdot)$ denotes the copula density function and the marginal density function is $f_i(x_i)$, $i=1,2,\cdots,n$. Then the multivariate joint distribution function and the corresponding PDF can be expressed as:

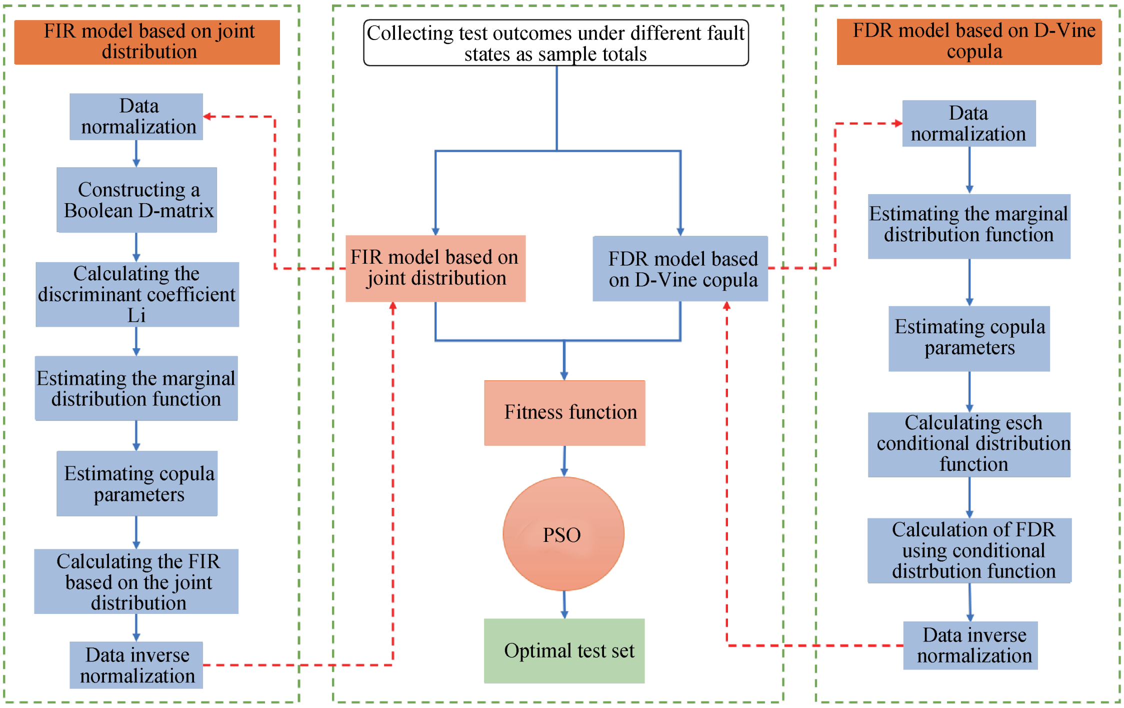


Fig.2 Test selection framework

$$F\left(x_1,x_2,\cdots,x_n\right)=C\left(F_1(x_1),F_2(x_2),\cdots,F_n(x_n)\right) \quad (5)$$

$$f(x_1,x_2,\cdots,x_n)=c\left(F_1(x_1),F_2(x_2),\cdots,F_n(x_n)\right)\cdot f_1(x_1)\cdot f_2(x_2)\cdots f_n(x_n) \quad (6)$$

### 3.1.2 FIR constraint model

For a particular fault $f_i$, successful fault isolation needs to be satisfied: 1) The row corresponding to $f_i$ must be different from all other rows in the fault-to-test correlation matrix; 2) The actual test results of $f_i$ should be consistent with the ideal test response. For condition 1), a similarity discriminant $LI_i$ is defined as ([16]):

$$LI_{ik}=\prod_{j=1}^{n}\left[\left(1-r_{ij}\right)\left(1-r_{kj}\right)+r_{ij}r_{kj}\right] \quad (7)$$

where $r_{ij}$ denotes the: value of fault $f_i$ under test point $t_j$ in the fault-test- correlation matrix, $r_{ij}=1$ when the test result $t_j^i$ is greater than the corresponding test threshold $thr_j$, otherwise $r_{ij}=0$.

$$LI_i=\prod_{k=1,k\neq i}^{m}\left(1-LI_{ik}\right) \quad (8)$$

If the test responses of $f_i$ and $f_k$ are identical, then $LI_{ik}=1$, $LI_i=0$; otherwise $LI_{ik}=0$, $LI_i=1$. Therefore, condition 1) is satisfied when $LI_i=1$.

For condition: 2), the actual test outcomes for $f_i$ should be consistent with the ideal test response, which can be expressed as $\cdot PI_i$:

$$PI_i=(<t_j^i\leq thr_j|r_{ij}=0\cap t_p^i\geq thp_p|r_{ip}=1>) \quad (9)$$

By adding a term $(-1)^{rij}$ to (10) to unify the form of the input variables, the FIR can be expressed as:

$$FIR(f_i)=LI_i\cdot P\left((-1)^{r_{ij}}t_j^i\leq(-1)^{r_{ij}}thr_j|j\in[1,n]\right)=LI_i\cdot F_X(x) \quad (10)$$

where the test outcomes of $f_i$ are represented by the $n$-dimensional random variable $X$, and the test thresholds are described by the specific upper limit vector $x$ in the joint distribution function.

Combining the above formulas while considering the selection vector $S$, the FIR based on the joint distribution can be finally obtained:

$$\begin{cases} FIR\left(f_i;S\right)=LI_i^*\cdot F_X(x^*) \\ LI_i^*=\prod_{k=1,i\neq i}^{m}\left(1-\prod_{j=1}^{n}\left[\left(1-r_{ij}\right)(1-r_{ij})+r_{ij}r_{kj}\right]^{*}\right) \\ X=(X_1,X_2,\cdots X_n),X_j=(-1)^{r_{ij}}t_j^i \\ x^*=(x_1^*,x_2^*,\cdots x_n^*),x_j^*=(-1)^{r_{ij}}thr_j^* \end{cases} \quad (11)$$

where

$$(-1)^{rij}thr_j^*\triangleq\begin{cases}(-1)^{rij}thr_j, & \text{if } s_j=1\\ +\infty, & \text{if } s_j=0\end{cases} \quad (12)$$

where the test selection vector is $S=(s_1,s_2,\cdots,s_n)$. If test $t_j$ is selected then $s_j=1$, otherwise $s_j=0$.

Combining (5), (11) and (12), the copula-based FIR model can be formulated as:

$$\begin{cases} FIR(f_i;S)=LI_i^*\cdot C(u_1,u_2,u_3,\cdots u_j;\theta_c),i\in[1,m] \\ LI_i^*=\prod_{k=1,k\neq i}^{m}\left(1-\prod_{j=1}^{n}\left[\left(1-r_j\right)\left(1-r_{kj}\right)+r_{ij}r_{kj}\right]^{s_j}\right) \\ u_i=F_X\left((-1)^{r_{ij}}thr_j^*\right) \\ X_j=(-1)^{r_{ij}}t_j^i, \qquad j\in[1,n] \end{cases} \quad (13)$$

There are several common copula functions, including Gaussian, *t*- , and Archimedean copulas. The Gaussian copula features a simple structure that enables the generation of numerous data samples with predefined statistical distributions and dependency information. As Gaussian distributions are widespread in real-world systems, this approach enhances the relevance of the modeling. In this paper, we employ the Gaussian copula function.

## 3.2 D-Vine copula-based FDR model

### 3.2.1 D-Vine copula theory for joint distribution

Multidimensional joint distributions, can be broken down into multiple two-dimensional copula functions. by decomposition methods, such as C-Vine copula and D-Vine copula. For C-Vine copula, one node (variable) per tree should be linked to each of the other nodes. This condition significantly restricts its applicability as the relevant multidimensional variables often lack primary variables. Conversely, the D-Vine copula offers structural flexibility and can mitigate the limitation caused by C-Vine. Hence, the D-Vine copula is selected to construct the FDR model in this paper. Fig. 3 illustrates a four-variable tree model based on D-Vine.

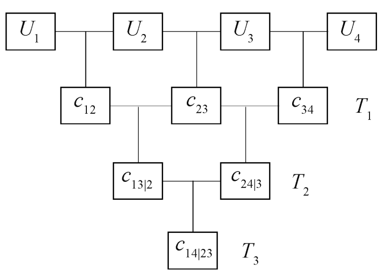


Fig.3 Four-variable D-Vine tree

$U=(U_1,U_2,U_3,U_4)$ is the marginal distribution function corresponding to the joint distribution $F_X(x)$, $X=(X_1,X_2,X_3,X_4)$. The PDF of the variable $X$ is calculated by D-Vine as follows:

$$\begin{aligned}
f(x_1,x_2,x_3,x_4) &= f(x_1)f(x_2)f(x_3)f(x_4)\cdot\\
&c_{12}(F_1(x_1),F_2(x_2))c_{23}(F_2(x_2),F_3(x_3))\\
&c_{34}(F_3(x_3),F_4(x_4))\cdot\\
&c_{1,312}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2))c_{2,4|3}\\
&(F_{2|3}(x_2|x_3),F_{4|3}(x_4|x_3))\cdot\\
&c_{1,4|2,3}(F_{1|2,3}(x_1|x_2,x_3),F_{4|2,3}(x_4|x_2,x_3))
\end{aligned} \tag{14}$$

For an $n$-dimensional D-Vine model containing $n-1$ layers of tree structures $T_j, j=1,2,3\cdots,n-1$, each tree $T_j$ has $n-j-1$ nodes and $n-j$ edges, with each edge representing a two-dimensional copula density function. The PDF of the multidimensional D-Vine model can then be expressed as follows:

$$\begin{aligned}
&f(x_1,x_2,\cdots,x_n)\\
&=\prod_{k=1}^{n} f_k(x_k)\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{i,i+j|i+1,\cdots,i+j-1}\\
&(F_{i|i+1,\cdots,i+j-1}(x_i|x_{i+1},\cdots,x_{i+j-1}),\\
&F_{i+j|i+1,\cdots,i+j-1}(x_{i+j}|x_{i+1},\cdots,x_{i+j-1}))
\end{aligned} \tag{15}$$

Then, combining Eq. (6) and Eq. (15) further yields:

$$\begin{aligned}
&c(F_1(x_1),F_2(x_2),\cdots,F_n(x_n))\\
&=\prod_{j=1}^{n-1}\prod_{i=1}^{n-j} c_{i,i+j|i+1,\cdots,i+j-1}\\
&(F_{i|i+1,\cdots,i+j-1}(x_i|x_{i+1},\cdots,x_{i+j-1}),\\
&F_{i+j|i+1,\cdots,i+j-1}(x_{i+j}|x_{i+1},\cdots,x_{i+j-1}))
\end{aligned} \tag{16}$$

where $F_{i|i+1,\cdots,i+j-1}(x_i|x_{i+1},\cdots,x_{i+j-1})$ is the conditional probability function and according to the conclusion given by Joe ([30]) the conditional probability of the joint distribution of multivariate variables is:

$$F(x|v)=\frac{\partial C_{x,v_j|v_{-j}}\left(F(x|v_{-j}),F(v_j|v_{-j})\right)}{\partial F(v_j|v_{-j})} \tag{17}$$

where $v$ is an n-dimensional vector, $v_j$ is any component of $v$, and $v_{-j}$ denotes the vector consisting of the parts of the vector $v$ that do not contain $v_j$ and $C(\cdot)$ represents the two-dimensional copula function under the D-Vine structure.

### 3.2.2 FDR constraint model

To account for the correlation among the outcomes of each test, an approach based on D-Vine copula is employed to construct the FDR, formulated as follows[15]:

$$F_X(x)=P(\bigcap_{j-1}^{n} t_j^i\le thr_j^*) \tag{18}$$

where $t_j^i$ denotes the test outcomes of $t_j$ when the $f_i$ occurs. If $t_j^i$ is larger than the corresponding test threshold $thr_j^*$ , then it can be drawn a conclusion that $t_j$ can detect the $f_i$, otherwise, the $t_j$ cannot detect the $f_i$.

If a given fault $f_i$ can be detected, it is assumed that at least one test among all test candidates can respond to the fault, and when considering which test point to use as an alternative test, the FDR can be expressed as follows:

$$FDR(f_i;S)=1-P\left(\left\{t_j^i\le thr_j^*[\,j\in[1,n]\right\}\right) \tag{19}$$

$$thr_j^*\triangleq\begin{cases}thr_j, \text{ if } s_j=1\\ +\infty, \text{ if } s_j=0\end{cases} \tag{20}$$

where, $t_j$ is selected then $s_j=1$, otherwise $s_j=0$.

Combining the above equations, the FDR based on the joint distribution is obtained:

$$\begin{cases}FDR(f_i;S)=1-F_X(x^*)\\ X=(X_1,X_2,\cdots X_n),X_j=t_j^i\\ x^*=(x_1^*,x_2^*,\cdots x_n^*),x_j^*=thr_j^*\end{cases} \tag{21}$$

In this section, the joint distribution $F_X(x^*)$ in Eq.

(21) can be solved for the FDR by utilizing the D-Vine structure introduced earlier in combination with Eq. (17). According to Bayes' theorem, the joint distribution is expressed as follows:

$$F(x_1,x_2,\cdots x_n)=F_1(x_1)F_{2|1}(x_2|x_1)F_{3|2,1}(x_3|x_2,x_1)\cdot \cdots F_{n|n-1,\cdots 1}(x_n|x_{n-1},\cdots x_1) \tag{22}$$

combining (17) and (22) yields:

$$F(x_1,x_2,\cdots x_n) = F_1(x_1)\frac{\partial C_{2,1}(F_2(x_2),F_1(x_1))}{\partial F_1(x_1)} \cdots \frac{\partial C_{n,v_j|v_{-j}}(F_{n|v_j}(x_n|x_{v_j}),F_{v_j|v_{-j}}(x_{v_j}|x_{v_{-j}}))}{\partial F_{v_j|v_{-j}}(x_{v_j}|x_{v_{-j}})} \tag{23}$$

From (17), (22) and (23) it is known that the joint distribution can be solved by decomposing it into multiple two-dimensional copulas using D-Vine copula. Thus, the joint distribution $F_X(x^*)$ in (21) can be similarly decomposed. The specific steps for resolving the FDR based on the D-Vine copula theory are as follows:

Step 1: The marginal distribution of each variable is obtained by:

$$u_1=F_1(x_1^*),u_2=F_1(x_2^*),\cdots u_n=F_n(x_n^*) \tag{24}$$

Step 2: According to (17), the joint conditional distribution of $x_2^*$ can be calculated by:

$$F_{2|1}(x_2^*|x_1^*)=\frac{\partial C_{2,1}(F_2(x_2^*),F(x_1^*);\theta_v)}{\partial F_1(x_1^*)} \tag{25}$$

Step 3: The joint conditional distributions of $x_3^*,x_4^*,\cdots x_n^*$ are achieved by:

$$F_{3|2,1}(x_3^*|x_2^*,x_1^*) = \frac{\partial C_{3,1|2}(F_{3|2}(x_3^*|x_2^*),F_{1|2}(x_1^*|x_2^*);\theta_v)}{\partial F_{1|2}(x_1^*(x_2^*)} \tag{26}$$

$$\vdots$$

$$F_{n|1,\cdots n-1}(x_n^*|x_1^*,\cdots x_{n-1}^*)= \frac{\partial C_{n,1|2,\cdots n-1}(F_{n|2,\cdots n-1}(x_n^*|x_2^*,\cdots x_{n-1}^*),F_{1|2,\cdots n-1}(x_1^*|x_2^*,\cdots x_{n-1}^*);\theta_v)}{\partial F_{1|2,\cdots n-1}(x_1^*|x_2^*,\cdots x_{n-1}^*)} \tag{27}$$

Then, the joint distribution $F_x(x^*)$ is expressed as follows:

$$F_X(x^*)=F_1(x_1^*)F_{2|1}(x_2^*|x_1^*)F_{3|2,1}(x_3^*|x_2^*,x_1^*) \cdots F_{n|1,\cdots n-1}(x_n^*|x_1^*,\cdots x_{n-1}^*) \tag{28}$$

### *3.3 Constraint Model Construction*

The constraint model for TSD is constructed by the weighted sum of FDR and FIR for each fault mode, as expressed below:

$$\begin{cases} FDR(F)=\dfrac{1}{1-p_0}\sum_{i=1}^{m}(p_iFDR(f_i)) \\ FIR(F)=\dfrac{1}{1-p_0}\sum_{i=1}^{m}(p_iFIR(f_i)) \end{cases} \tag{29}$$

where $p_i$ represents the prior probability of the corresponding $f_i$ and $\sum_{i=0}^{m}p_i=1$.

The TSD goal is to select the optimal test set that minimizes the required test cost under all test points. The function of total testing cost, which is also known as the objective function, can be expressed as:

$$C=\sum_{j=1}^{n}ct_j\cdot s_j \tag{30}$$

Based on (13), (21), (28), (29), and (30), the constraint model for overall test selection is constructed as:

$$\begin{gathered} \min C=\sum_{j=1}^{n}ct_j\cdot s_j \\ s.t. FDR(F)\ge\overline{FDR} \\ FIR(F)\ge\overline{FIR} \\ j\in[1,n] \end{gathered} \tag{31}$$

where $\overline{FDR}$ and $\overline{FIR}$ are the smallest values that satisfy *FDR* and *FIR*.

# 4 Test selection optimization

### *4.1 Copula Function Parameter Estimation*

From equations (13), (25), (26) and (27), it can be seen that there are unknown parameters $\theta_c$ and $\theta_v$ in the copula function during the construction of FIR and FDR, and the parameters reflect the correction between the variables in the copula function. The MLE method is used to estimate $\theta_c$ and $\theta_v$ based on the sample data. The logarithmic likelihood function of FIR is introduced to evaluate the parameter $\theta_c$ as:

$$\begin{cases} \ln L_c(\cdot)=\prod_{q=1}^{k}\left[\ln c(u_{q1},u_{q2},\cdots,u_{qj};\theta_c)\right] \\ u_j=F_X((-1)^{r_{ij}}thr_j^*) \\ X_j=(-1)^{r_{ij}}t_j^i, \qquad j\in[1,n] \end{cases} \tag{32}$$

where $u_{qj}$ is the $q$ th sample data of $u_j$.

For the FDR construction process, the multi-dimensional joint distribution is decomposed into multiple two-dimensional copulas. Logarithmic likelihood function is given to estimate the parameter $\theta_c$ as:

$$\ln L_v(\cdot)= \sum_{q=1}^{k}\left[\ln c_{i,1|2,\cdots,i-1}\begin{pmatrix} F_{qi|q2,\cdots,q(i-1)}(x_{qi}|x_{q2},\cdots,x_{q(i-1)}), \\ F_{1|q2,\cdots,q(i-1)}(x_1|x_{q2},\cdots,x_{q(i-1)});\theta_v \end{pmatrix}\right], \quad i\in[3,n] \tag{33}$$

In particular, if $i=2$, the copula function is already a two-dimensional form and no further decomposition is needed. Finally, select the maximum $L_c(\cdot)$ and $L_v(\cdot)$ to estimate $\theta_c$ and $\theta_v$.

### 4.2 PSO based optimal test selection

As previously mentioned, selecting the optimal test set must also ensure compliance with the constraints of FDR and FIR. The PSO method offers strong global search capabilities and fast convergence. Compared to other methods, PSO has significant advantages in test point optimization. Unlike genetic algorithms' complex crossover and mutation operations in, PSO features a simpler updating mechanism and typically converges to a better solution more quickly. In contrast to simulated annealing algorithms, PSO does not require gradual cooling to converge, resulting in lower computational overhead and reducing the risk of getting stuck in a local optimum. Additionally, compared to differential evolution algorithms, PSO is easier to implement and does not rely on frequent differential operations, making it less costly in parameter tuning. Consequently, PSO has played a key role in the optimal selection of measurement points in recent years.

In this section, the PSO algorithm will be used to select the best test. All particles in PSO represent attributes of the solution space, with each particle's fitness generated based on its current position to reflect its quality. Each particle possesses a velocity vector, and its position is represented by binary values of 1 or 0. Greater particle velocities indicate higher probabilities of selection. In the specific application of this paper, each particle represents a test point selection scheme, where a value of 0 or 1 for one of the dimensions indicates whether the corresponding test point is selected or not, corresponding to $s_j$ in the text. The fitness is used to evaluate the advantages and disadvantages of each test point selection scheme, the fitness in this paper refers to the cost of testing, which corresponds to Eq. (30), and the particle with the smallest fitness needs to be selected. The specific algorithm is outlined in Algorithm 1.

## 5 Case study

Negative feedback circuits (NFC) are widely used in analog electronics, finding application across various electronic systems such as marine fuel injection systems, intelligent manufacturing systems, and photovoltaic power generation systems [24]. These circuits play a key role in feeding a voltage or current input signal and feeding it back to the output part of the circuit at a controlled ratio. The schematic diagram of the NFC employed in this paper is depicted in Fig. 4.

### 5.1 Platform Description

In the NFC shown in Fig. 4, the input $V_1$ is a sine wave signal with frequency 1khz and amplitude 7mv, and the supply voltage $V_{cc}$ is 15V. Here, 6 test points, $T=\{t_1,t_2,t_3,t_4,t_5,t_6\}$ are used to collect data. Five types of faults are applied for experimental verification as shown in Table 1.

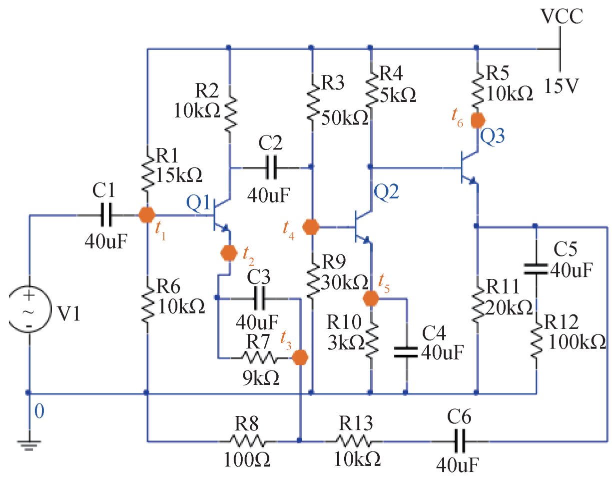


Fig.4 NFC schematic diagram

**Algorithm 1** Procedure of PSO

1. A set of particles in $n$-dimensional space ($n$-dimension is the number of test points $t_j$) {1, 1, 1, ....... ,1}, {1, 1, 1, ...... ,0}, ……. Are generate corresponds to the current position 'xbest', and each particle consists of 0 and 1, reflecting whether the corresponding test point is selected or not.

2. The fitness $fx$ is initialized. Here $fx$ is the test cost per particle, i. e., $fx=C=\sum_{j=1}^{n} ct_j \cdot s_j$. The initial fitness $fx$ is larger than the test cost when all test points are selected.

3. The $FDR(F)$ and $FIR(F)$ for each particle as well as the fitness $fx$ corresponding to the particle are calculated based on Eq. (13) , Eq. (21), Eq. (28) and Eq. (29). For particles that do not satisfy the conditions $FDR(F) \geq \overline{FDR}$ and $FIR(F) \geq \overline{FIR}$, the corresponding $fx$ is replaced by the initial one.

4. Compare the current $fx$ of each particle with the initial $fx$. If the current $fx$ is less than the initial $fx$, then the initial $fx$ is replaced by the current one.

5. Update the velocity vector and position vector of each particle

6. Implement a catastrophe strategy: escape from the current best position and regenerate $m$ random particles. Repeat steps 3-4 until a termination criterion is satisfied (the maximum number of iterations is reached).

7. Find the smallest $fx$ among all particles, whose location 'xbest' is the global best location 'gbest' and the optimal combination of test points.

Table 1 Information fault types

| Fault type | Fault mode |
|---|---|
| $f_1$ | $Q_1\ B-C\ short$ |
| $f_2$ | $Q_1\ C-E\ short$ |
| $f_3$ | $Q_2\ C-E\ short$ |
| $f_4$ | $R_8=130\Omega$ |
| $f_5$ | $Q_1\ C-E\ short \& R_9=39K\Omega$ |

Totally, $10^3$ test data are collected for each fault type via Monte Carlo method from six test points. The data are assumed to follow a normal distribution. A threshold is established as a criterion to determine whether the test responds to a fault or not. The collected test data in the state of no-fault $f_0$ are used to obtain the mean value $\mu$ and the standard deviation $\sigma$. The $\mu+3\sigma$ is set as the judgment threshold, which is set as: $thr$=(6.22,5.60,0.064, 5.58,4.93,12.12).

Based on the simulation data, the ideal value (mean value) for each fault state can be calculated, as shown in Table 2.

Table 2 Ideal values of test results under different faults

| | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $f_0$ | 5.97 | 5.36 | 0.05 | 5.30 | 0.05 | 11.74 |
| $f_1$ | 6.82 | 6.20 | 0.06 | 5.37 | 4.72 | 11.74 |
| $f_2$ | 6.00 | 7.10 | 0.07 | 5.36 | 4.72 | 11.74 |
| $f_3$ | 5.97 | 5.30 | 0.06 | 5.60 | 5.62 | 12.51 |
| $f_4$ | 5.97 | 5.35 | 0.07 | 5.30 | 4.72 | 11.74 |
| $f_5$ | 6.00 | 7.15 | 0.07 | 6.10 | 5.47 | 12.34 |

Then the ideal value of each fault-to-test result is compared with the threshold $thr_j$. If the ideal value of fault $f_i$ obtained under test $t_j$ is larger than the corresponding threshold value, it indicates that test $t_j$ can detect fault $f_i$, and the corresponding position is marked as 1; otherwise, the corresponding position is marked as 0. A fault-to-test correlation matrix, as illustrated in Table 3, is obtained by comparisons for all fault-to-test.

Table 3 Fault-to-test correlation matrix

| | $t_1$ | $t_2$ | $t_3$ | $t_4$ | $t_5$ | $t_6$ |
|---|---|---|---|---|---|---|
| $f_0$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $f_1$ | 1 | 1 | 0 | 0 | 0 | 0 |
| $f_2$ | 0 | 1 | 1 | 0 | 0 | 0 |
| $f_3$ | 0 | 0 | 0 | 1 | 1 | 1 |
| $f_4$ | 0 | 0 | 1 | 0 | 0 | 0 |
| $f_5$ | 0 | 1 | 1 | 1 | 1 | 1 |

## 5.2 TSD Model Construction and Comparisons

Two different approaches Bernoulli distribution (BD), and joint distribution (JD), are utilized to validate the effectiveness of the proposed model based on D-Vine copula theory (DV).

Since {0,0,0,0,0,0}, {0,0,0,0,0,1},……,{1,0,0,0,0,0} are not relevant, they are not included in the alternative test sets. While {0, 0, 0, 0, 1, 1}, …… , {1, 1, 1, 1, 1, 1} are chosen to verify the correlation issue in TSD modeling, totaling 57 sets. The verification are performed for each of the three methods and experimental results are shown in Fig.5-Fig. 9.

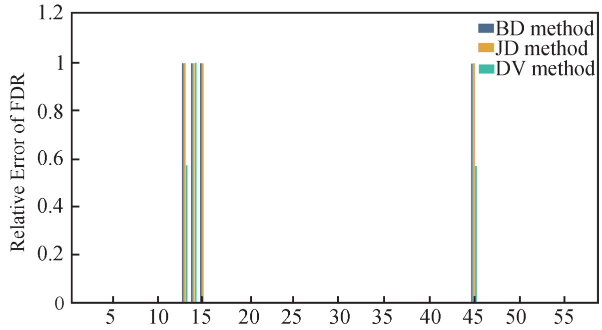


Fig.5 Comparison results of FDR under fault $f_1$

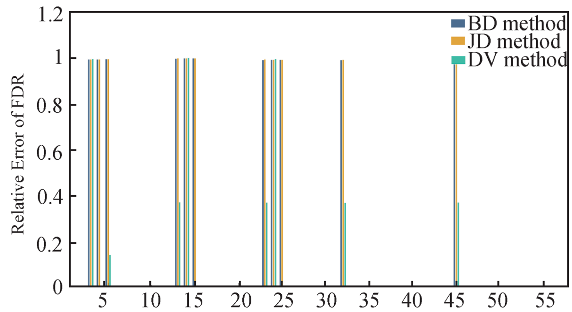


Fig.6 Comparison results of FDR under fault $f_2$

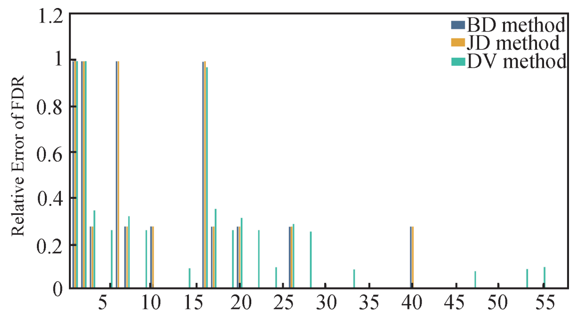


Fig.7 Comparison results of FDR under fault $f_3$

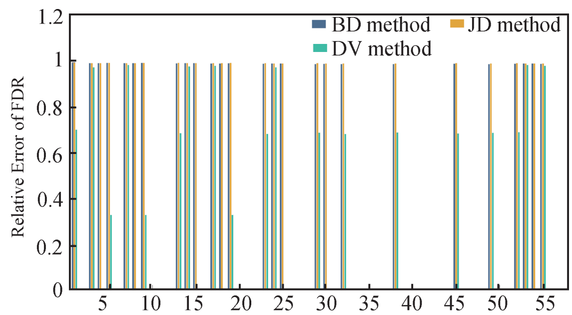


Fig.8 Comparison results of FDR under fault $f_4$

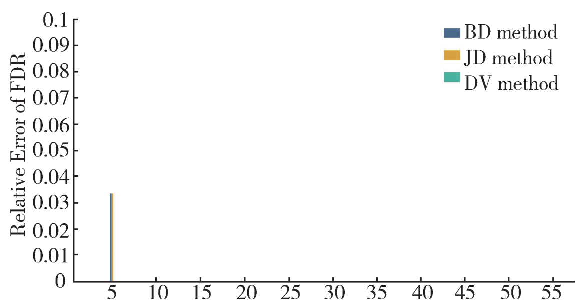


Fig.9 Comparison results of FDR under fault $f_5$

The relative errors of the FDR for faults $f_1$ through $f_5$ across various constraint models are shown in Fig 5 to 9, with lower error values indicating better performance. These figures illustrate that the FDR error values based on the BD, JD, and DV models generally perform well in most test subsets. However, in a few specific subsets, such as {1, 0, 0, 0, 0, 1, 0}, {1, 0, 0, 0, 0, 1}, and {0, 1, 0, 0, 0, 0, 1} under fault $f_4$, the FDR error values for the BD and JD models are noticeably higher. In contrast, the DV model achieves excellent results in these cases. To facilitate a comprehensive comparison of the actual results, Table 4 presents the average and maximum values of the actual FDR.

Table 4 Mean and maximum values of FDR

| | Average FDR | | | Maximum FDR | | |
|---|---|---|---|---|---|---|
| | BD | JD | DV | BD | JD | DV |
| $f_1$ | 0.9296 | 0.9297 | 0.9659 | 1 | 1 | 1 |
| $f_2$ | 0.8083 | 0.8090 | 0.9251 | 1 | 1 | 1 |
| $f_3$ | 0.8963 | 0.8963 | 0.8876 | 1 | 1 | 1 |
| $f_4$ | 0.5459 | 0.5461 | 0.7649 | 0.9999 | 0.9999 | 1 |
| $f_5$ | 0.9994 | 0.9994 | 1 | 1 | 1 | 1 |

From Table 4, it is apparent that the average FDR of the model constructed based on DV is slightly lower than the other two methods for fault $f_3$. However, the maximum FDR value of the DV-based model under this fault is nearly equal to the other two methods. For the remaining four faults, the DV-based FDR model exhibits a notable advantage, with its average FDR value significantly outperforming the other two methods, while maintaining well performance in terms of the maximum FDR value. This indicates the existence of a larger subset of tests with higher FDRs, presenting more opportunities to select test points with equal or superior performance at lower costs. In summary, the FDR model based on D-Vine copula theory emerges as a preferred choice.

The proposed FIR constraint model is also compared between BD and JD-based methods, which are given from Fig. 10 to Fig. 14. These figures clearly illustrate that the JD-based FIR model, in most of the test subsets, is better than the BD-based FIR model. And the superiority of the JD model is more prominently

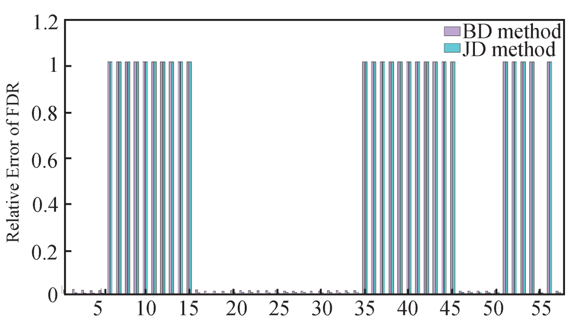


Fig.10 Comparison results of FIR under fault $f_1$

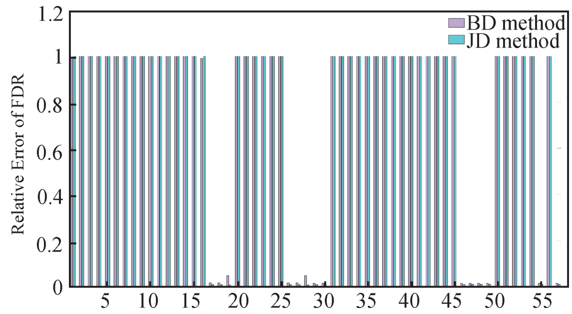


Fig.11 Comparison results of FIR under fault $f_2$

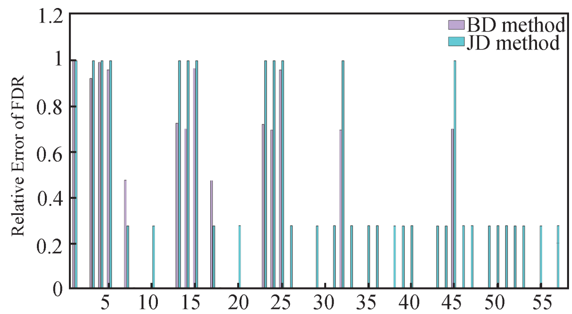


Fig.12 Comparison results of FIR under fault $f_3$

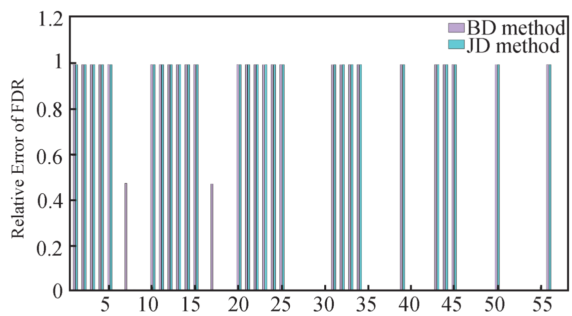


Fig.13 Comparison results of FIR under fault $f_4$

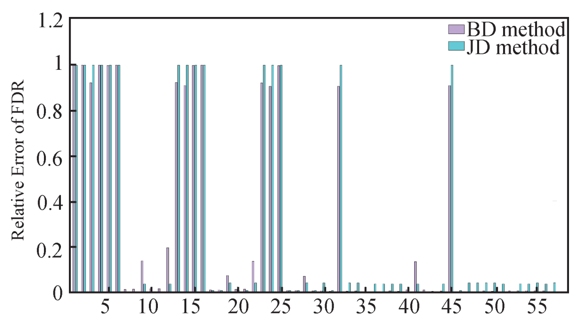


Fig.14 Comparison results of FIR under fault $f_5$

demonstrated, especially in the fault $f_5$. However, in a few subsets, for example, the subsets {0, 1, 0, 1, 0, 0}, {0, 1, 0, 0, 1, 0} and {0, 1, 0, 0, 0, 1} under faults $f_3$ and $f_5$, the JD-based FIR model is less effective than the BD-based FIR model. The mean and maximum values of the actual FIR are listed in Table 5 to further compare the experimental results:

From Table 5, it is apparent that the average FIR of the model constructed based on JD is slightly lower than the BD method for fault $f_2$. However, the maximum FIR value of the JD-based model under this fault is nearly

Table 5 Mean and maximum values of FIR

| | Average FIR | | Maximum FIR | |
|---|---|---|---|---|
| | BD | JD | BD | JD |
| $f_1$ | 0.5377 | 0.5419 | 0.9899 | 1 |
| $f_2$ | 0.2771 | 0.2473 | 0.9874 | 0.9938 |
| $f_3$ | 0.1617 | 0.6721 | 0.3068 | 0.9992 |
| $f_4$ | 0.5060 | 0.5245 | 0.9975 | 0.9987 |
| $f_5$ | 0.0468 | 0.7208 | 0.0941 | 0.9999 |

equal to the BD method. For the remaining four faults, the JD-based FIR model exhibits a notable advantage, with its average FIR value significantly outperforming the BD method, while maintaining a good performance in terms of the maximum FIR value. In summary, the FIR model based on JD theory emerges as a preferred choice.

### *5.3 Optimal Test Set Selection*

While many test points are needed to gather fault information, there is often significant redundancy in these tests. Therefore, optimizing the selection of test points is essential to reduce testing and maintenance costs. The PSO algorithm is utilized to achieve the optimal selection of the test set at a minimum test cost, thereby reducing the cost of the equipment over its entire life cycle. In the PSO algorithm, the population consists of 57 test subsets, with a maximum iteration count of 100. The FDR and FIR minimum criteria are set to (0.9, 0.9) and the test costs are randomly generated between 0.1 and 1. The proposed DV-PSO algorithm is compared with the BD-PSO and JD-PSO algorithm to illustrate the superiority. The optimal test sets and test costs for these methods are list in Table 6. Additionally, the fitness curves for the respective methods are provided.

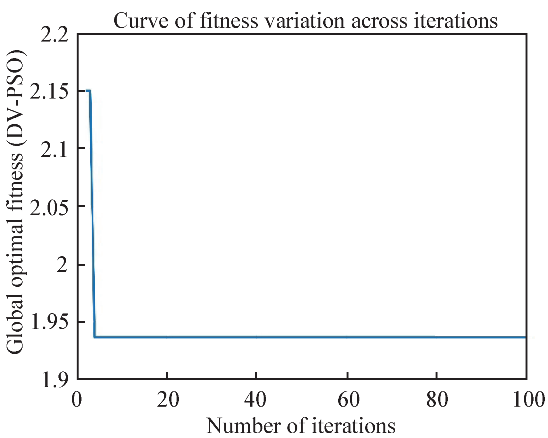


Fig.15 Fitness curve under DV-PSO

From Table 6, the DV-PSO algorithm can choose 3 test points with a test cost of 1.9363, while the JD-PSO algorithm and BD-PSO algorithm choose 4 test points with a test cost of 2.1506. Thus, the DV-PSO algorithm has the least number of test points and the smallest test cost when both FDR and FIR satisfy the constraints. As, the proposed method results in a reduction in the testing cost, it demonstrates the economy and applicability of the proposed method in the test selection design.

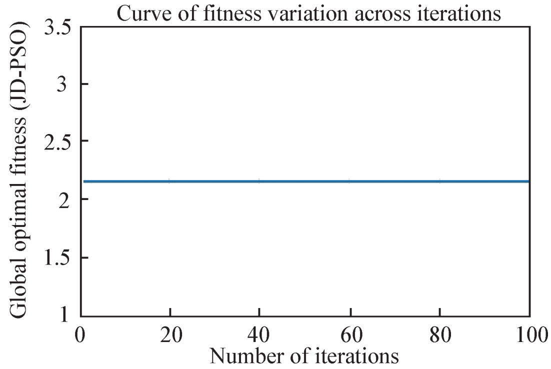


Fig.16 Fitness curve under JD-PSO

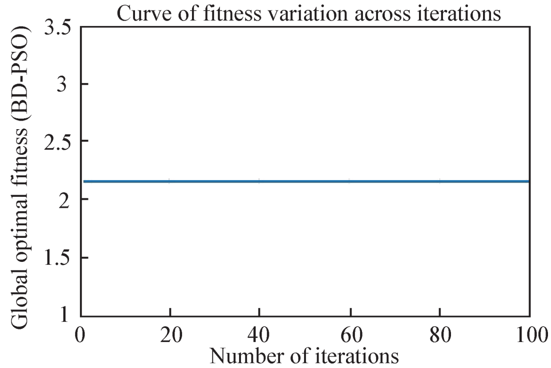


Fig.17 Fitness curve under BD-PSO

Table 6 Test selection comparison results

| | test set | test costs |
|---|---|---|
| DV-PSO | $t_1, t_2, t_6$ | 1.9363 |
| JD-PSO | $t_1, t_2, t_3, t_6$ | 2.1506 |
| BD-PSO | $t_1, t_2, t_3, t_6$ | 2.1506 |

# 6 Conclusions

This study proposes a new method for modeling FIR and FDR by copula and D-Vine copula to capture the correlations between test outcomes. Additionally, the PSO algorithm is employed for optimal test point selection. The effectiveness of the proposed method is validated compared to traditional methods such as BD and JD through experiments on NFC.

However, this study focuses on modeling testability metrics for single faults, and it is effective for a limited number of multiple faults. In real systems, numerous multi-fault scenarios often lead to excessively ambiguous relationships, resulting in relatively inaccurate outcomes. Future research will use D-Vine copula and copula methods to address fuzzy issues, enabling accurate modeling and diagnosis of faults in complex systems.

## Author Contribution:

Xiuli Wang: Supervision, Project administration, Writing-review & editing. Dongdong Xie: Methodology, Analysis and interpretation of data, Writing-original draft, Software. Yejian Gong: Conceptualization, Writing-review & editing, Supervision. Yang Li: Supervision, Conceptualization, Project administration, Writing-review & editing. Chun Liu: Writing review & editing, Supervision. Defeng He: Writing review & editing, Supervision.

## Foundation Information:

This work was supported by the National Natural Science Foundation of China (No. 62303293, 62303414), the China Postdoctoral Science Foundation (No. 2023M732176, 2023M741821) and the Zhejiang Province Postdoctoral Selected Foundation (No. ZJ2023143).

## Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

## Conflicts of Interest:

The authors declare no competing interests.

## Dates:

Received 22 August 2024; Accepted 4 February 2025; Published online 31 March 2025.

# References

[1] Wang J, Sun X, Zhang C, et al. "An integrated methodology for system-level early fault detection and isolation." *Expert Systems with Applications* 201 (**2022**): 117080.

[2] Su F, Liu C, Stratigopoulos H G. "Testability and dependability of AI hardware: Survey, trends, challenges, and perspectives." *IEEE Design & Test 40.2* (**2023**): 8-58.

[3] Kang H, Ma H. "Fault detection and isolation of actuator failures in jet engines using adaptive dynamic programming." *Applied Mathematics and Computation* 414 (**2022**): 126664.

[4] Wang H, Liu Z, Nunez A, et al. "Entropy-based local irregularity detection for high-speed railway catenaries with frequent inspections." *IEEE Transactions on Instrumentation and Measurement 68.10* (**2018**): 3536-3547.

[5] Liu Z, Liu K, Zhong J, et al. "A high-precision positioning approach for catenary support components with multiscale difference." *IEEE Transactions on Instrumentation and Measurement 69.3* (**2019**): 700-711.

[6] Ni T, Wen X, Amrouch H, et al. "Introduction to the Special Issue on Design for Testability and Reliability of Security-aware Hardware." *ACM Transactions on Design Automation of Electronic Systems 29.1* (**2023**): 1-3.

[7] Li M, Lin Y, Gupta S. "Design for testability (dft) for rsfq circuits." 2023 IEEE 41st VLSI Test Symposium (VTS). IEEE, **2023**.

[8] Cui W, Xue W, Li L, et al. "A review on testability verification test and fault injection technology." **2020** Global Reliability and Prognostics and Health Management (PHM-Shanghai) (**2020**): 1-5.

[9] Su F, Liu C, Stratigopoulos H G. Stratigopoulos. "*Special Issue on Testability and Dependability of Artificial Intelligence Hardware." IEEE Design & Test 40.2* (**2023**): 5-7.

[10] Schennach S M, Wilhelm D. "A simple parametric model selection test." *Journal of the American Statistical Association 112.520* (**2017**): 1663-1674.

[11] Cui Y, Shi J, Wang Z. "Intermittent failure process and false alarm interaction modelling of threshold-based monitoring built-in tests (BITs)." *International Journal of Production Research 54.6* (**2016**): 1610-1626.

[12] Shi J, Li J, Shi M. "Method of automated BIT false alarms simulation based on EDA." *2011 6th IEEE Conference on Industrial Electronics and Applications. IEEE* , **2011**.

[13] Aslam M S, Ullah R, Dai X, et al "Event-triggered scheme for fault detection and isolation of non-linear system with time-varying delay." *IET Control Theory & Applications* 14.16 (**2020**): 2429-2438.

[14] Lu N, Jiang B, Meng X, et al. "*Diagnosis, diagnosticability analysis, and test point design for multiple faults based on multisignal modeling and blind source separation* ." IEEE Transactions on Systems, Man, and Cybernetics: Systems 50.1 (**2017**): 137-148.

[15] Ye X, Chen C, Kang M, et al. "A joint distribution-based testability metric estimation model for unreliable tests." *IEEE Access* 6 (**2018**): 42566-42577.

[16] Zhang S, Pattipati K R, Hu Z, et al. "*Optimal selection of imperfect tests for fault detection and isolation* ." IEEE Transactions on Systems, Man, and Cybernetics: Systems 43.6 (**2013**): 1370-1384.

[17] Di P, Wang X, Chen T, et al. "Multisensor data fusion in testability evaluation of equipment." *Mathematical Problems in Engineering* **2020**.1 (2020): 7821070.

[18] Garousi V, Felderer M, Kılıçaslan F N. "A survey on software testability." Information and Software Technology 108 (**2019**): 35-64.

[19] Shi J, He Q, Wang Z. "Integrated Stateflow-based simulation modelling and testability evaluation for electronic built-in-test (BIT) systems." *Reliability Engineering & System Safety* 202 (**2020**): 107066.

[20] Li Y, Wang X, Lu N, et al. "Conditional joint distribution-based test selection for fault detection and isolation." *IEEE Transactions on Cybernetics 52.12* (**2021**): 13168-13180.

[21] Li Y, Zio E, Lu N, et al. "Joint distribution-based test selection for fault detection and isolation under multiple faults condition." *IEEE Transactions on Instrumentation and Measurement* 70(**2020**): 1-13.

[22] Tang X, Xu A, Niu S. "KKCV-GA-based method for optimal analog test point selection." *IEEE Transactions on*

*Instrumentation and Measurement 66.1* (**2016**): 24-32.

[23] Tang X, Xu A, Li R, et al. "Simulation-based diagnostic model for automatic testability analysis of analog circuits." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 37.7* (**2017**): 1483-1493.

[24] Li Y, Chen H, Lu N, et al. "Data-driven optimal test selection design for fault detection and isolation based on CCVKL method and PSO." *IEEE Transactions on Instrumentation and Measurement* 71 (**2022**): 1-10.

[25] Lewandowski D, Kurowicka D, Joe H. "Generating random correlation matrices based on vines and extended onion method." *Journal of multivariate analysis 100.9* (**2009**): 1989-2001.

[26] Aas K, Czado C, Frigessi A, et al. "*Pair-copula constructions of multiple dependence* ." Insurance: Mathematics and economics 44.2 (**2009**): 182-198.

[27] Jiang C, Zhang W, Han X, et al. "A vine-copula-based reliability analysis method for structures with multidimensional correlation." *Journal of Mechanical Design 137.6* (**2015**): 061405.

[28] Li Q, Cai Y, Wang H, et al. "An efficient D-vine copula-based coupling uncertainty analysis for variable-stiffness composites." *Composite Structures* 219 (**2019**): 221-241.

[29] Czado C, Nagler T. "Vine copula based modeling." *Annual Review of Statistics and Its Application 9.1* (**2022**): 453-477.

[30] Joe H. "Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters." *Lecture notes-monograph series* (**1996**): 120-141.