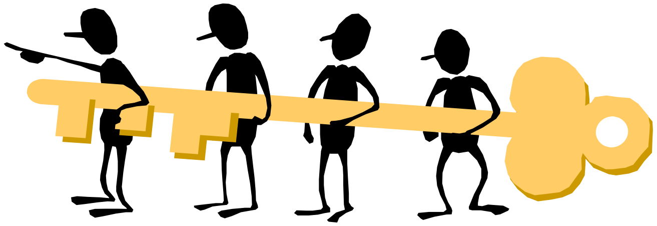
Introduction to Mathematics

Farokhlagha Moazami
Cyberspace Research Institute
Shahid Beheshti University
Tehran, Iran
f_moazemi@sbu.ac.ir



Outline

- Galois field
- Homomorphism and isomorphism
- Hypotheses test
- T-test

Galois Field:

Fields with a finite number of elements are called Finite Fields or Galois Fields and abbreviated as GF.

When p is prime, the residue class $\mathbb{Z}_p$ is a field that is represented as $GF(p)$, and commonly referred to as the prime field.

When $p = 2$, this field $GF(2)$ is called the binary field and is popular for fast and efficient implementations.

Galois Field:

Let f be a polynomial in $\mathbb{Z}_p$, of degree $n,$ the polynomial is **irreducible**, if it cannot be factored into polynomials, g and h which are polynomials in $\mathbb{Z}_p[x]$ of positive degree.

The irreducible polynomials can be imagined to correspond to prime numbers in the domain of integers.

Galois Field:

Example: The polynomial $f(x) = x^2 + x + 1$ is irreducible in $\mathbb{Z}_2[x]$.
Assume $f(x) = g(x)h(x)$. Since

$$\deg\big(f(x)\big) = \deg\big(h(x)\big) + \deg\big(g(x)\big)$$

then

$$\deg\big(g(x)\big) = \deg\big(h(x)\big) = 1 \text{ in } \mathbb{Z}_2[x].$$

Thus $f(0) = 0$ or $f(1) = 0$, but $f(0) \neq 0$ and $f(1) \neq 0$

Hence, the polynomial $f(x)$ is irreducible.

Galois Field:

Theorem: For a non-constant polynomials $f(x) \in \mathbb{Z}_p[x]$, the ring $\frac{\mathbb{Z}_p[x]}{<f(x)>}$ is a field if and only if $f(x)$ is irreducible in $\mathbb{Z}_p[x]$.

Consider a field K and $f(x)$ irreducible over K. Then, $L = \frac{K[x]}{<f(x)>}$ is a field extension.

$GF(p^n)$ is an extension of $GF(p)$.

Extension of a Field:

Consider $GF(2^2) = \frac{GF(2)[x]}{<f(x)>}$,

where $f(x) = x^2 + x + 1$ is an irreducible polynomial in $GF(2)[x]$.

$GF(2^2)$ is an extension field of $GF(2)$.

Theorem: Every non-constant **polynomial** over a field **has** a **root** in some **extension field**.

$GF(2^m)$ is known as binary extension finite fields or binary finite fields.

$GF(2^m)$

Advantages:

- Modern computer systems are built on the binary number system.
- With m bits all possible elements of $GF(2^m)$ can be represented.
- The simple hardware required for computation of some of the commonly used arithmetic operations such as addition and squaring.

Addition in binary extension fields can be easily performed by a simple XOR. There is no carry generated.

Squaring in this field is a linear operation and can also be done using XOR circuits.

$GF(2^m)$

The polynomial

$$a(x) = a_0 + a_1 x + \cdots + a_{m-1} x^{m-1} + a_m x^m$$

is a polynomial over $GF(2)$ if

$$a_0, a_1, \ \cdots, a_{m-1}, a_m \in GF(2).$$

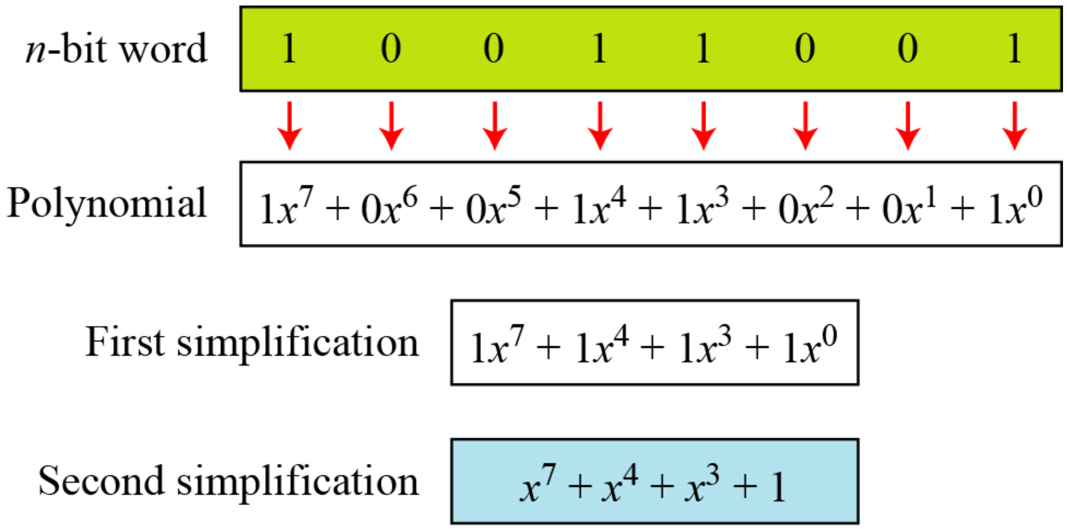
Let $f(x)$ be an **irreducible polynomial** of **degree m** over $GF(2)$,then

$$GF(2^m) = \frac{GF(2)}{<f(x)>}.$$

Hence, all elements of $GF(2^m)$ can be represented by polynomials of degree $m - 1$ over $GF(2)$.

Representation of element of $GF(2^m)$

we can represent the 8-bit word (10011001) using a polynomials



n-bit word: 1 0 0 1 1 0 0 1

Polynomial: $1x^7 + 0x^6 + 0x^5 + 1x^4 + 1x^3 + 0x^2 + 0x^1 + 1x^0$

First simplification: $1x^7 + 1x^4 + 1x^3 + 1x^0$

Second simplification: $x^7 + x^4 + x^3 + 1$

and vice versa the polynomial $x^5 + x^2 + x$ can be represented by the word 00100110

Since $n = 8$, the expanded polynomial is

$$0x^7 + 0x^6 + 1x^5 + 0x^4 + 0x^3 + 1x^2 + 1x^1 + 0x^0 \longrightarrow 00100110$$

Addition in $GF(2^m)$

Let $a(x), b(x) \in GF(2^m)$ be denoted by

$$a(x) = \sum_{i=0}^{m-1} a_i x^i \qquad b(x) = \sum_{i=0}^{m-1} b_i x^i,$$

$$a(x) + b(x) = \sum_{i=0}^{m-1} (a_i + b_i) x^i.$$

Here the + between

Addition in $\mathrm{GF}(2^{\mathrm{m}})$ is XOR

Example: Let a … hen

$a(x) + b(x) = (x$ … $+ x^2 + 1$ in $GF(2^8)$

Representation of $x^5 + x^2 + x$ is 00100110 and of $x^3 + x^2 + 1$ is 00001101 so

$$\begin{array}{r} 00100110 \\ \oplus \quad 00001101 \\ 00101011 \end{array}$$

Multiplication in $GF(2^m)$

Multiplication is **not** as trivial as addition or squaring.

The product of the two polynomials $a(x)$ and $b(x)$ is given by

$$a(x).b(x) = \sum_{i=0}^{n-1} b(x)a_i x^i \quad mod(p(x))$$

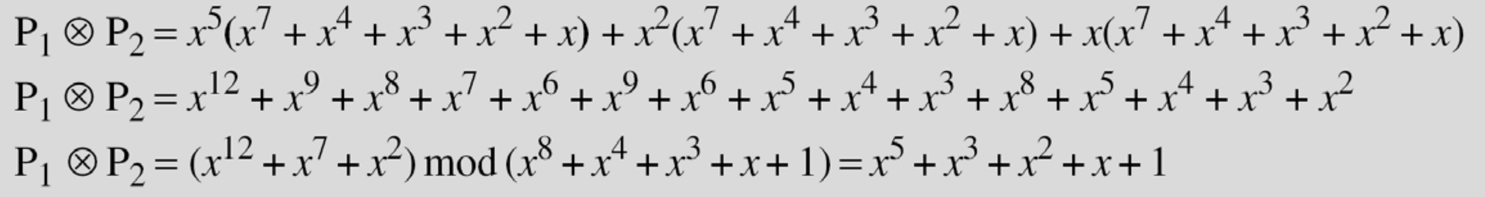
Most multiplication algorithms are $O(n^2)$.

Inversion is the most complex of all field operations.

Even the best technique to implement inversion is several times more complex than multiplication.

Multiplication in $GF(2^m)$

Example: Let $P_1 = x^5 + x^2 + x$ and $P_2 = x^7 + x^4 + x^3 + x^2 + x$ find $P_1 \times P_2$ in $GF(2^8)$ with **irreducible** polynomial $\boldsymbol{x^8 + x^4 + x^3 + x + 1}$.

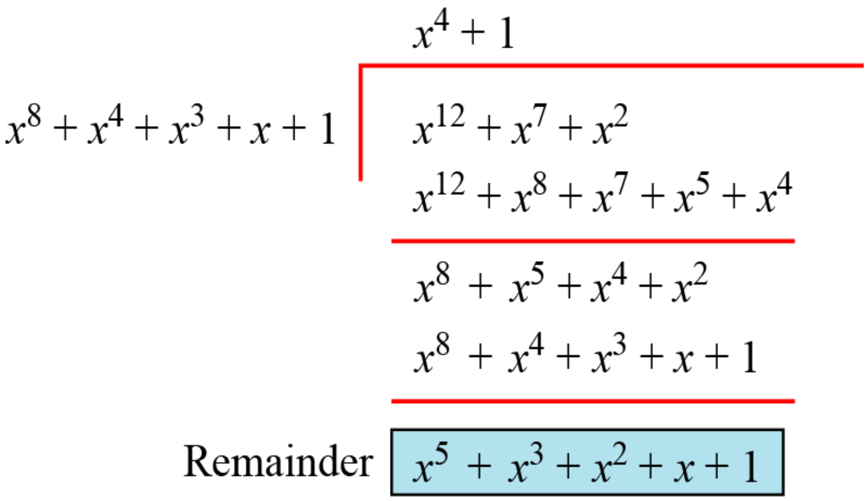


$$P_1 \otimes P_2 = x^5(x^7 + x^4 + x^3 + x^2 + x) + x^2(x^7 + x^4 + x^3 + x^2 + x) + x(x^7 + x^4 + x^3 + x^2 + x)$$

$$P_1 \otimes P_2 = x^{12} + x^9 + x^8 + x^7 + x^6 + x^9 + x^6 + x^5 + x^4 + x^3 + x^8 + x^5 + x^4 + x^3 + x^2$$

$$P_1 \otimes P_2 = (x^{12} + x^7 + x^2) \bmod (x^8 + x^4 + x^3 + x + 1) = x^5 + x^3 + x^2 + x + 1$$

Divide the polynomial of degree 12 by the polynomial of degree 8 (the modulus) and keep only the remainder



$x^4 + 1$

$x^8 + x^4 + x^3 + x + 1$

$x^{12} + x^7 + x^2$

$x^{12} + x^8 + x^7 + x^5 + x^4$

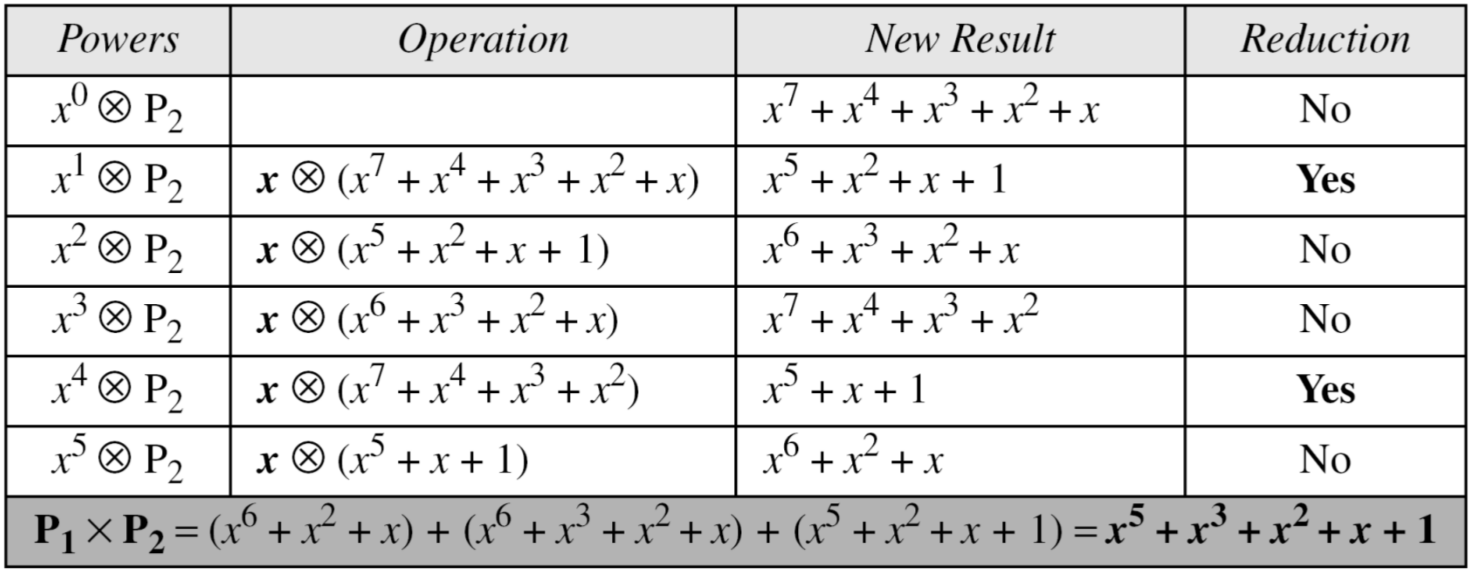
$x^8 + x^5 + x^4 + x^2$

$x^8 + x^4 + x^3 + x + 1$

Remainder $\boxed{x^5 + x^3 + x^2 + x + 1}$

Multiplication in $GF(2^m)$

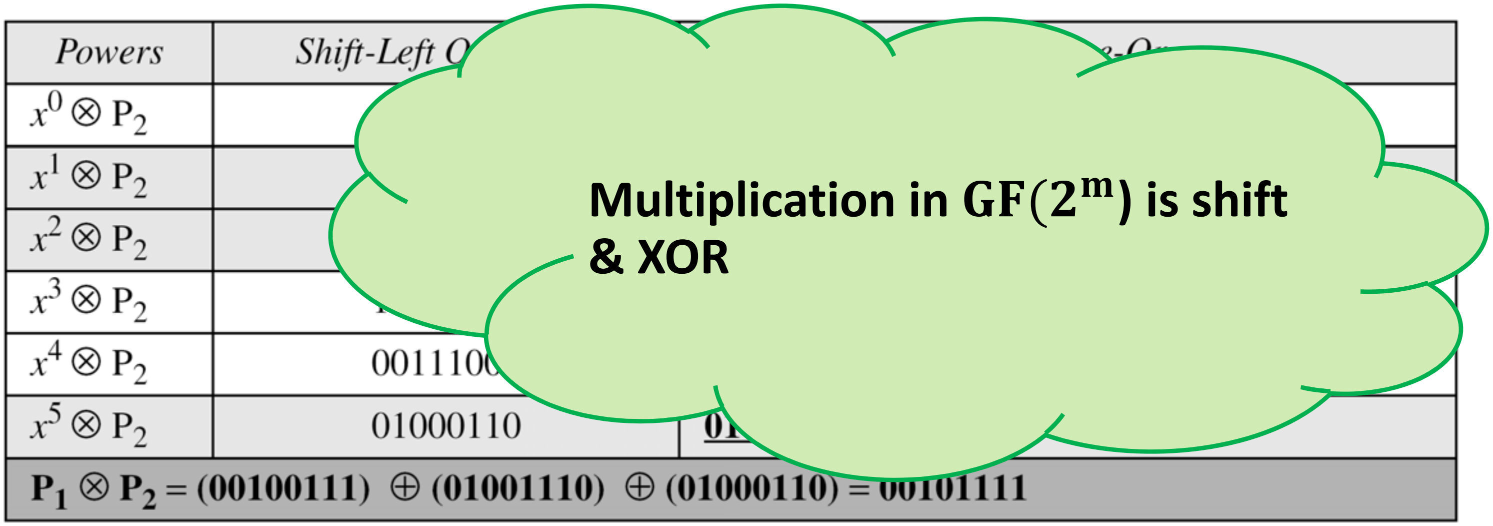
We first find the partial result of multiplying x^0, x^1, x^2, x^3, x^4 and $x^5$by P_2. Note that each calculation depends on the previous result.



Powers	*Operation*	*New Result*	*Reduction*
$x^0 \otimes \mathrm{P}_2$		$x^7 + x^4 + x^3 + x^2 + x$	No
$x^1 \otimes \mathrm{P}_2$	$\boldsymbol{x} \otimes (x^7 + x^4 + x^3 + x^2 + x)$	$x^5 + x^2 + x + 1$	**Yes**
$x^2 \otimes \mathrm{P}_2$	$\boldsymbol{x} \otimes (x^5 + x^2 + x + 1)$	$x^6 + x^3 + x^2 + x$	No
$x^3 \otimes \mathrm{P}_2$	$\boldsymbol{x} \otimes (x^6 + x^3 + x^2 + x)$	$x^7 + x^4 + x^3 + x^2$	No
$x^4 \otimes \mathrm{P}_2$	$\boldsymbol{x} \otimes (x^7 + x^4 + x^3 + x^2)$	$x^5 + x + 1$	**Yes**
$x^5 \otimes \mathrm{P}_2$	$\boldsymbol{x} \otimes (x^5 + x + 1)$	$x^6 + x^2 + x$	No
$\mathbf{P_1} \times \mathbf{P_2} = (x^6 + x^2 + x) + (x^6 + x^3 + x^2 + x) + (x^5 + x^2 + x + 1) = \boldsymbol{x^5 + x^3 + x^2 + x + 1}$			

Multiplication in $GF(2^m)$

We have $P_1 = 000100110$, $P_2 = 10011110,$ modulus = 100011010 (nine bits). We show the exclusive or operation by $\oplus$.



Powers	Shift-Left O	
$x^0 \otimes P_2$		
$x^1 \otimes P_2$		
$x^2 \otimes P_2$		
$x^3 \otimes P_2$		
$x^4 \otimes P_2$	0011100	
$x^5 \otimes P_2$	01000110	01
$P_1 \otimes P_2 = (00100111) \oplus (01001110) \oplus (01000110) = 00101111$		

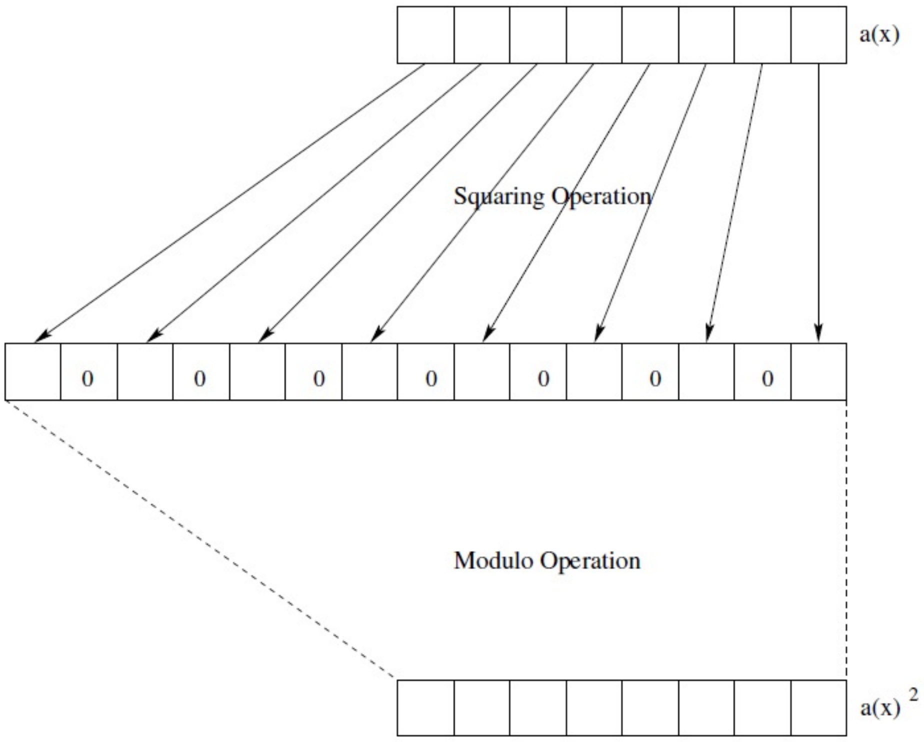
Multiplication in $\mathbf{GF(2^m)}$ is shift & XOR

Square operation in $GF(2^m)$

The square of the polynomial $a(x) \in GF(2^m)$ is given by

$$a(x)^2 = \sum_{i=0}^{m-1} a_i x^{2i} \quad mod(p(x)).$$

The squaring essentially spreads out the input bits by inserting zeroes in between two bits.



Squaring Circuit

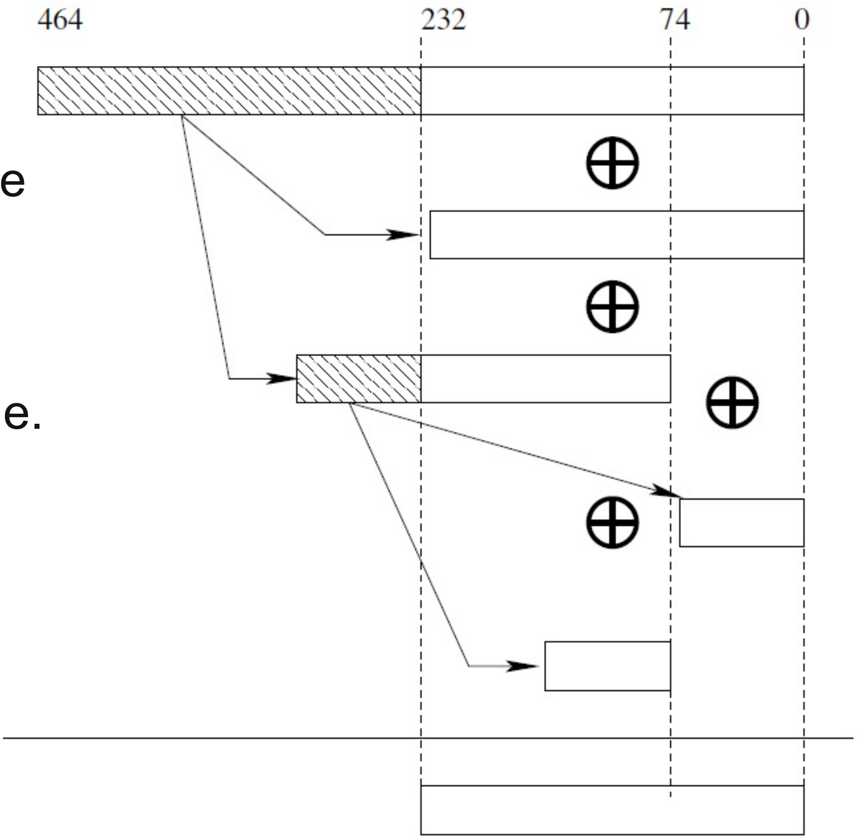
Modular operation in $GF(2^m)$

The modular operation is the remainder produced when divided by the field's irreducible polynomial.

If a certain class of irreducible polynomials is used, the modular operation can be easily done.

Consider the irreducible trinomial $x^m + x^n + 1$, having a root α and $1 < n < \frac{m}{2}$. Therefore

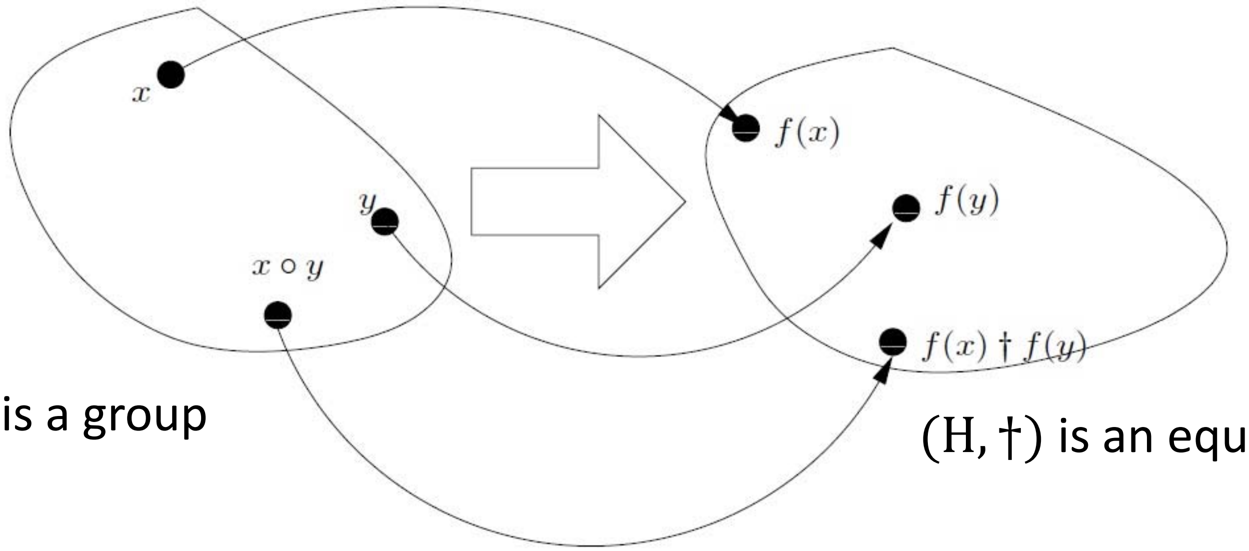
$$\begin{aligned}
\alpha^m + \alpha^n + 1 &= 0, \\
\alpha^m &= \alpha^n + 1, \\
\alpha^{m+1} &= \alpha^{n+1} + \alpha, \\
&\vdots \\
\alpha^{2m-3} &= \alpha^{n+m-3} + \alpha^{m-3}, \\
\alpha^{2m-2} &= \alpha^{n+m-2} + \alpha^{m-2}.
\end{aligned}$$



Modular Reduction with Trinomial $x^{233} + x^{74} + 1$

Homomorphism

A group, ring or a field can be expressed in several equivalent forms.



(G, o) is a group

$(\mathrm{H}, \dagger)$ is an equivalent form of (G, o)

For two groups, (G, o) and $(\mathrm{H}, \dagger)$, a surjective function $f\colon G \rightarrow H$ is said to be a homomorphism if and only if: $f(x\mathrm{o}y) = f(x) \dagger f(y)$.

Properties of Homomorphic Group

Theorem: If $f: G \to H$ is a group homomorphism then $f(e_1) = e_2$, where e_1 is the identity of G and e_2 is the identity of H.

Proof:…

Theorem: If $f: G \to H$ is a group homomorphism then for every $x \in G, f(x^{-1}) = f(x)^{-1}$.

Proof:….

An injective (one-to-one) homomorphism is called an isomorphism.

Definition: Let $(\boldsymbol{R_1}, +, \text{o})$ and $(\boldsymbol{R_2}, +', \text{o}')$ be rings and consider a surjective function, $f: R_1 \to R_2$. It is called a ring isomorphism if and only if:

- ➢ $f(a+b) = f(a) +' f(b)$ for every a and b in R_1.
- ➢ $f(a\text{o}b) = f(a)\text{o}' f(b)$ for every a and b in R_1.

Properties:

1) $f(0) = 0$ and $f(-x) = -f(x)$ for every $x \in R_1$.

2) $f(1) = 1'$ where 1 and $1'$ are multiplicative identities of R_1 and R_2, respectively.

3) If x is a unit in R_1, then $f(x)$ is a unit in R_2, and $f(x^{-1}) = f(x)^{-1}$.

These also holds for fields.

Application:

The isomorphism is utilized to transform a given field into another isomorphic field

Perform operations in this field

Then **transform back** the solutions.

The operations in the newer field are more **efficient** to implement than the initial field.

Definition: The pair of the fields $GF(2^n)$ and $GF(2^n)^m$are called a composite field, if there exists irreducible polynomials, $Q(Y\ )$ of degree n and $P(X)$ of degree m, which are used to extend $GF(2)$ to $GF(2^n)$, and $GF(2^n)^m$ from $GF(2^n)$.

A composite field is isomorphic to the field, $GF(2^k)$, where $k = m \times n$.

Example: Consider the fields $GF(2^4)$, elements of which are the following 16 polynomials with binary coefficients:

$\mathbf{0}$	z^2	z^3	$z^3 + z^2$
1	$z^2 + 1$	$z^3 + 1$	$z^3 + z^2 + 1$
z	$z^2 + z$	$z^3 + z$	$z^3 + z^2 + z$
$z + 1$	$z^2 + z + 1$	$z^3 + z + 1$	$z^3 + z^2 + z + 1$

Irreducible polynomials of degree 4:

$f_1(z) = z^4 + z + 1, f_2(z) = z^4 + z^3 + 1,\ f_3(z) = z^4 + z^3 + z^2 + z + 1.$

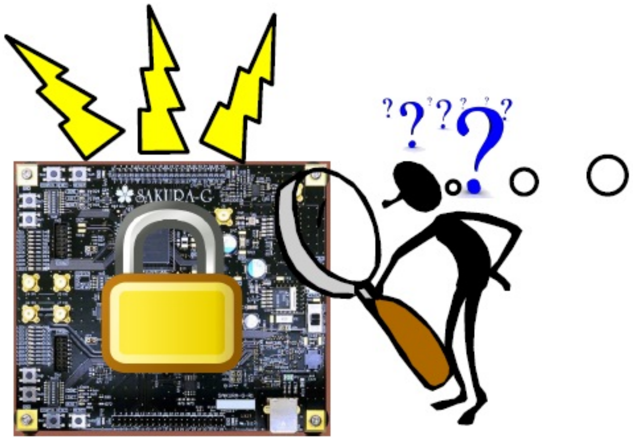
The resulting fields, F_1, F_2, F_3 **all** have the **same** elements.

But the **operations** are **different** for example consider $z.z^3$
Which is $z + 1$ in F_1,
$z^3 + 1$ in F_2 and
is $z^3 + z^2 + z + 1$ in F_3

The fields are isomorphic.

Hypothesis Testing

Motivation



Does the chip leak information?

Problem: Evaluation is not trivial.

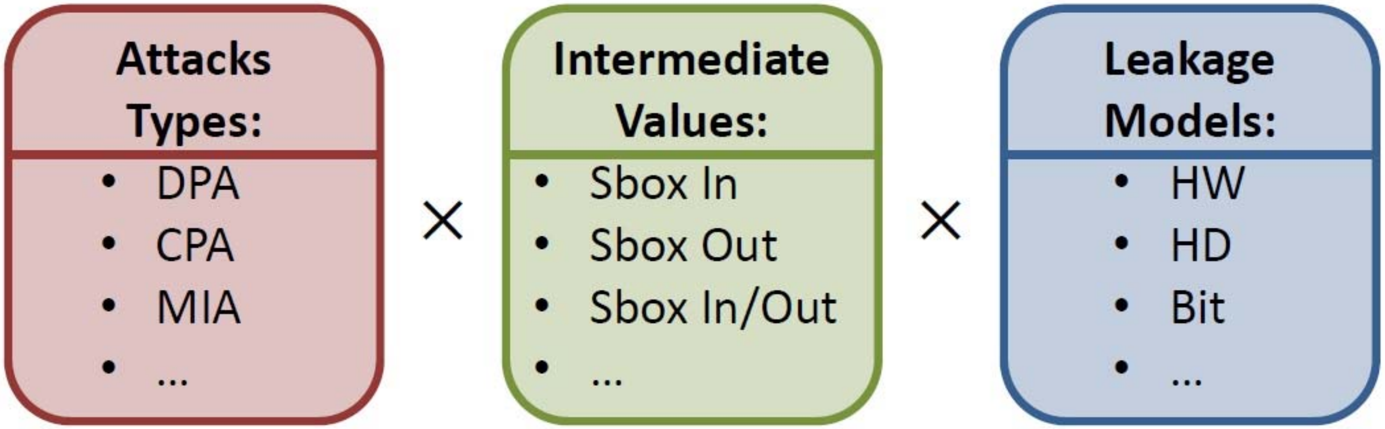
NIST *Non-Invasive Attack Testing Workshop, 2011*

Goal: Establish testing methodology capable of robustly assessing the physical vulnerability of cryptographic devices.

This slide is courtesy of **Tobias Schneider**

Motivation

Perform state-of-the-art attacks on the device under test (DUT)



Attacks Types:

- DPA
- CPA
- MIA
- …

×

Intermediate Values:

- Sbox In
- Sbox Out
- Sbox In/Out
- …

×

Leakage Models:

- HW
- HD
- Bit
- …

Problems:

- High computational complexity
- Requires lot of expertise
- Does not cover all possible attack vectors

This slide is courtesy of **Tobias Schneider**

Motivation

Standardization bodies intend to establish a leakage assessment methodology. One of such proposals is the t-test that is able to relax the dependency between the evaluations and the device's underlying architecture.

Advantages:

- Independent of architecture
- Independent of attack model
- Fast & simple

Problems:

- No information about hardness of attack
- Possible false positives if no care about evaluation setup

Question:

Whether two sets of data are significantly different from each other?

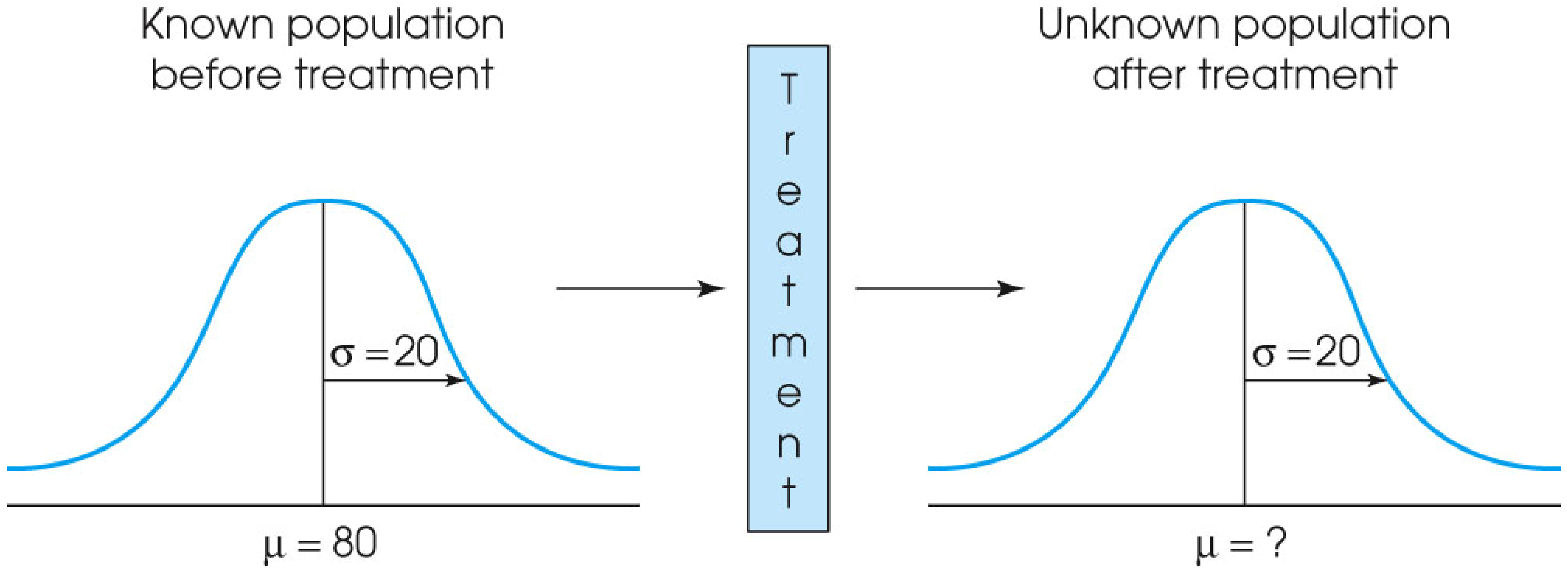
Hypothesis Testing

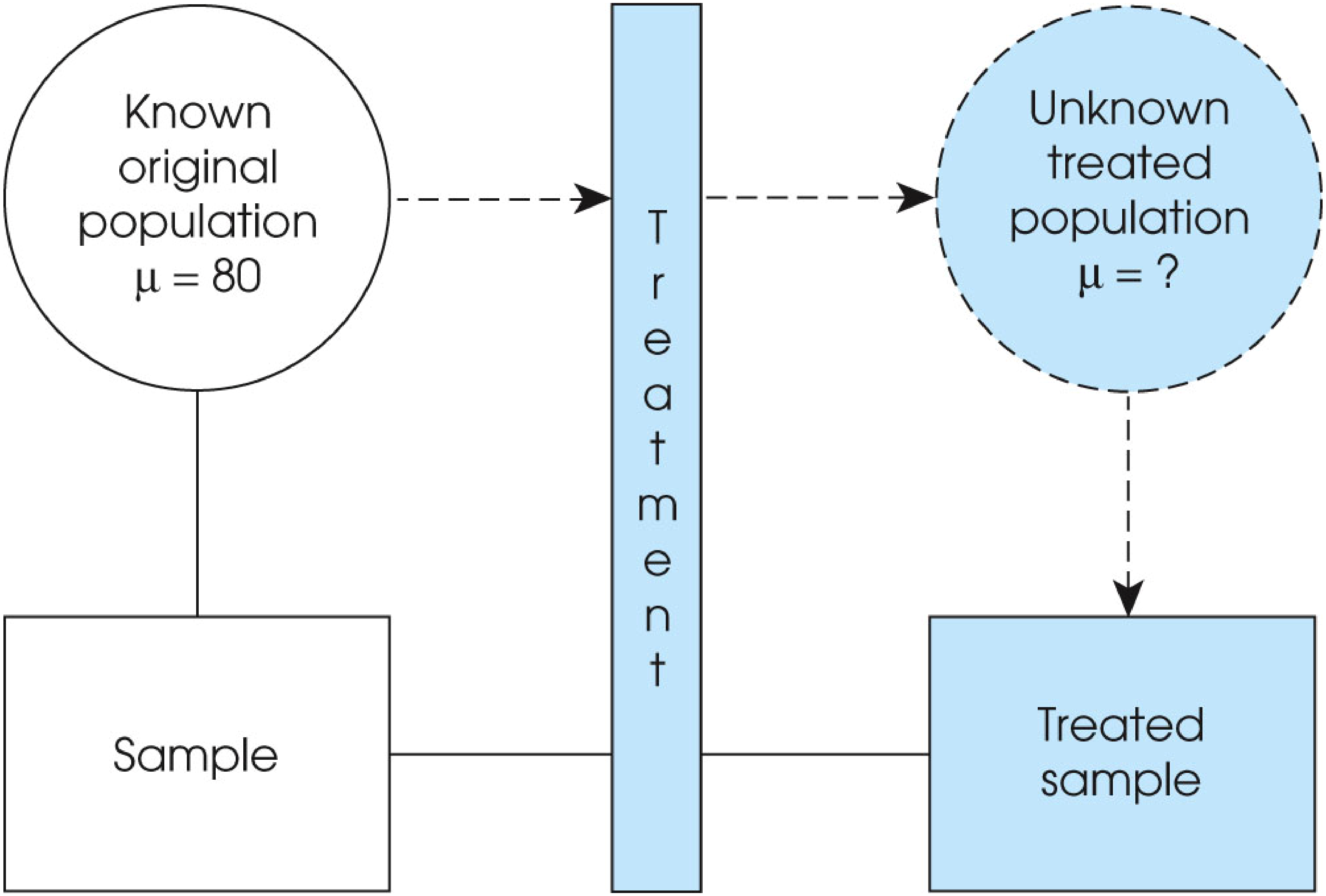
The general goal of a hypothesis test is to rule out chance (sampling error) as a plausible explanation for the results from a research study.

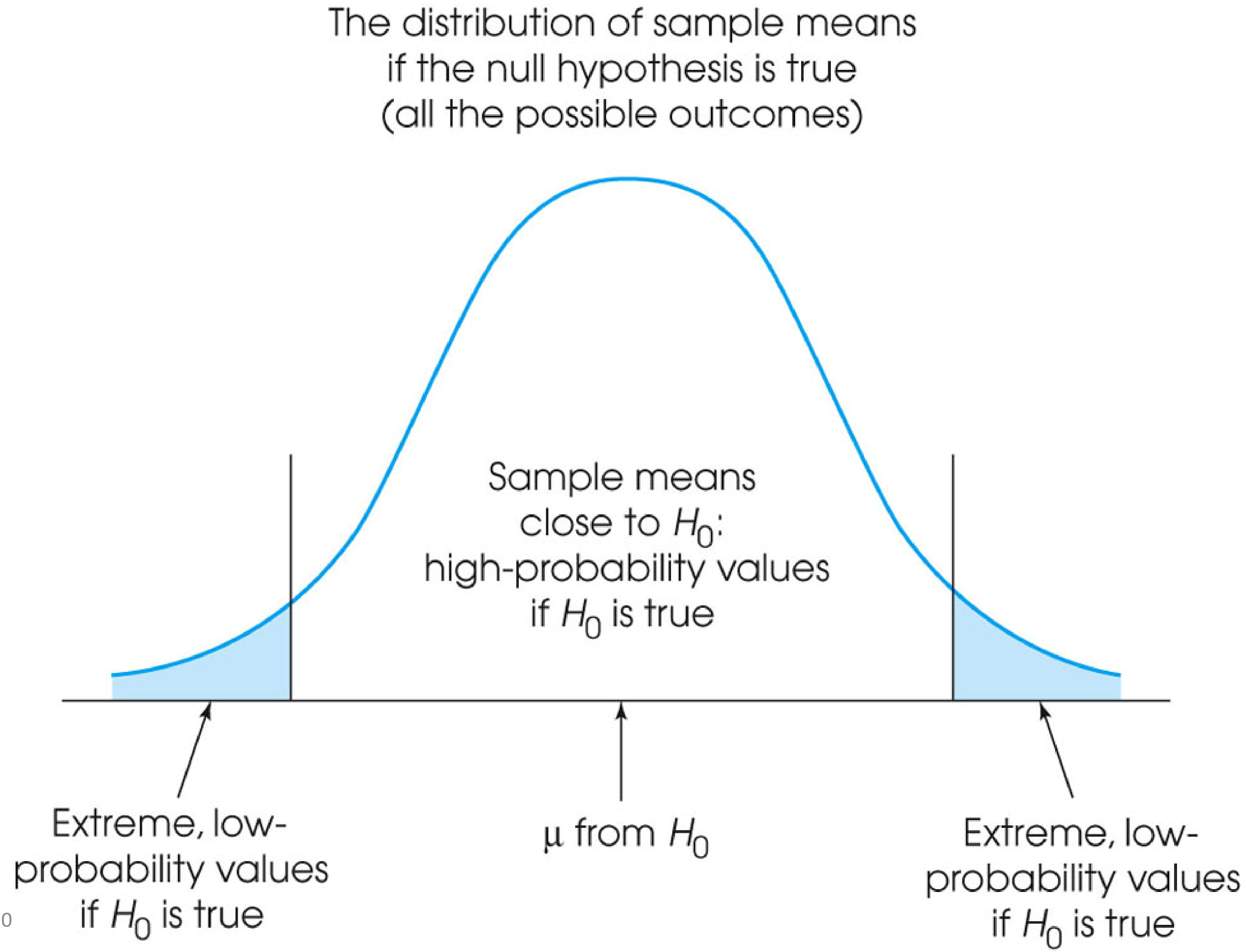
Hypothesis testing is a technique to help determine whether a specific treatment has an effect on the individuals in a population.

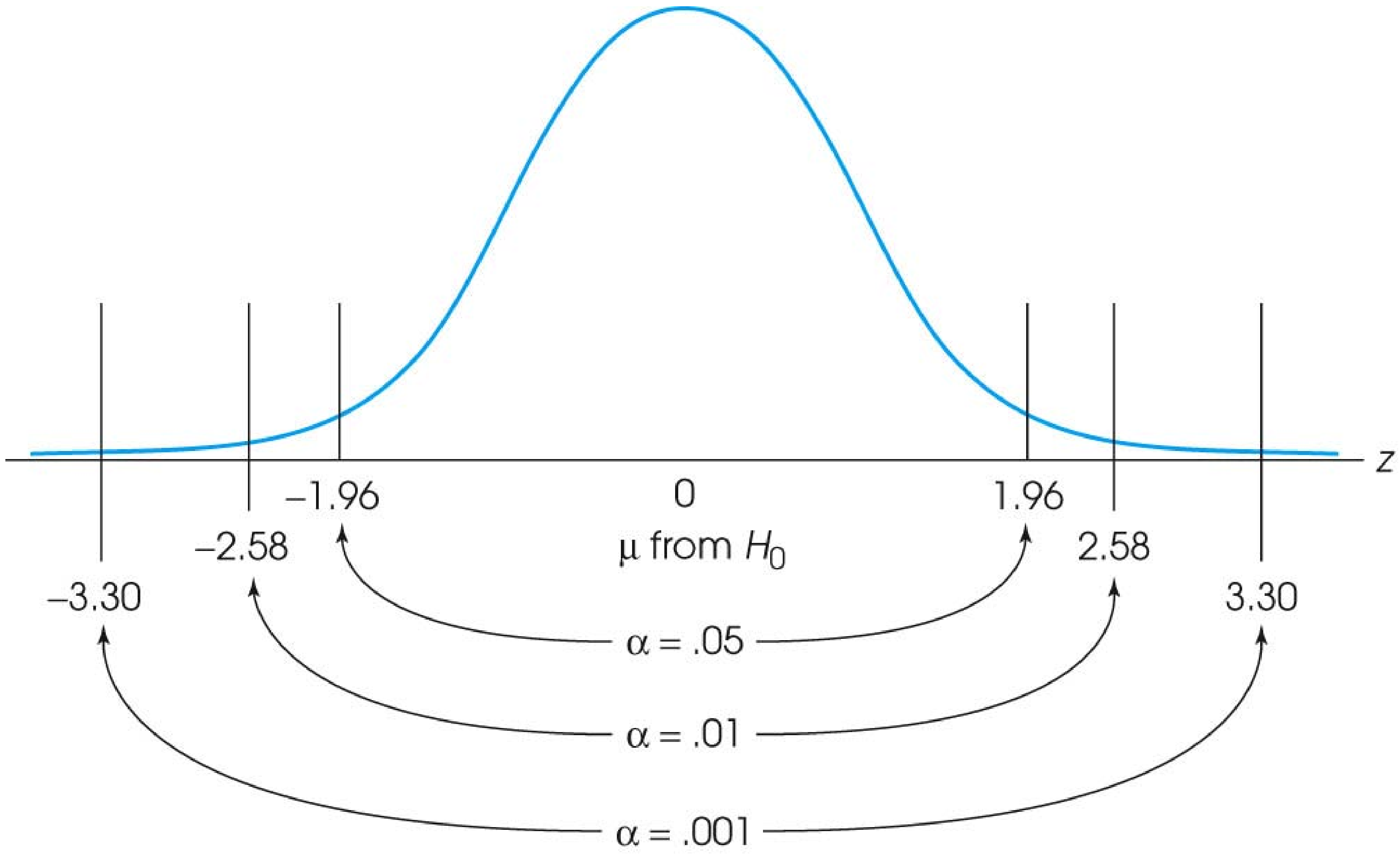
If the individuals in the sample are noticeably different from the individuals in the original population, we have evidence that the treatment has an effect.

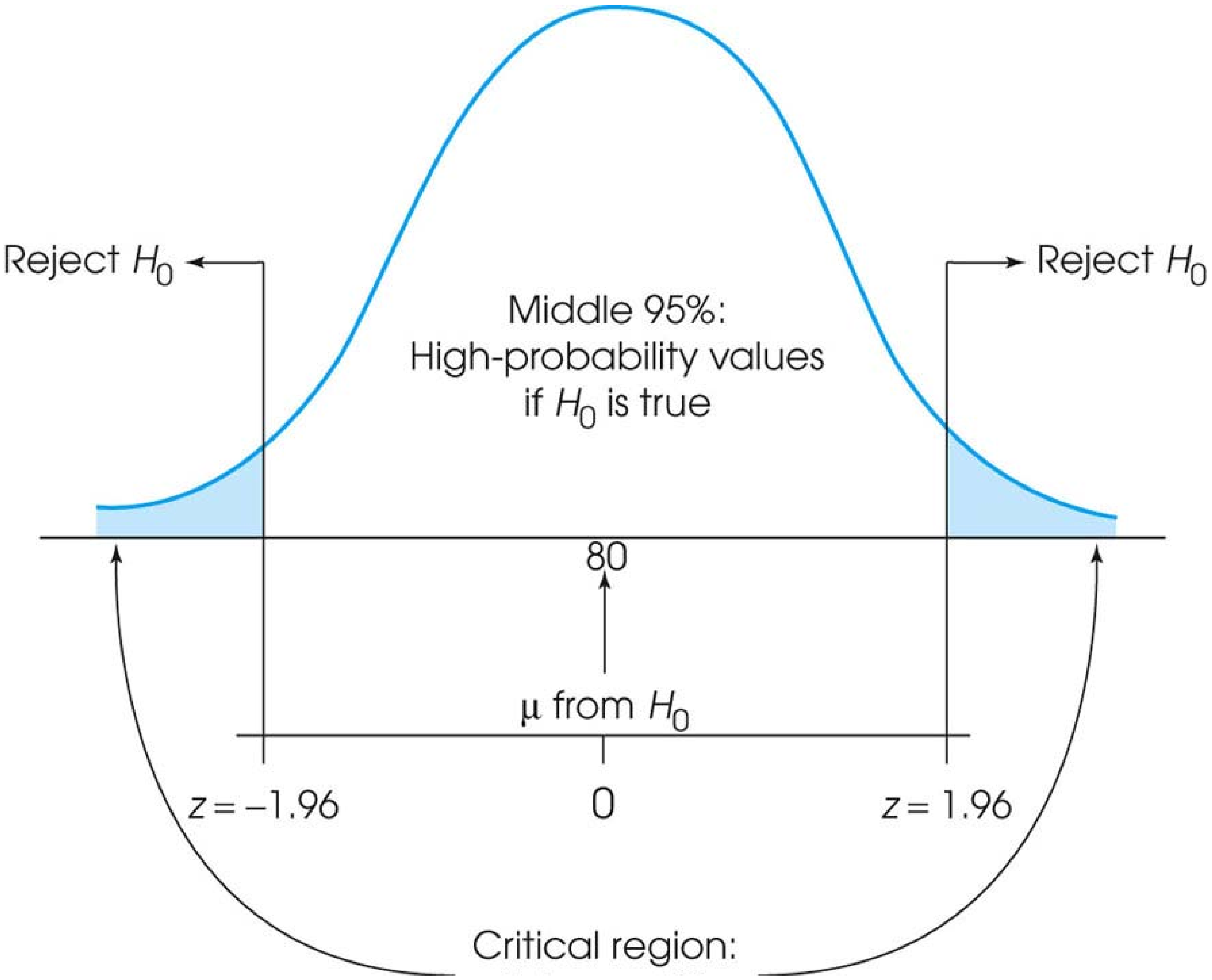
However, it is also possible that the difference between the sample and the population is simply sampling error


Known population
before treatment
σ = 20
μ = 80
Treatment
Unknown population
after treatment
σ = 20
μ = ?


Known original population $\mu = 80$
Treatment
Unknown treated population $\mu = ?$
Sample
Treated sample


The distribution of sample means
if the null hypothesis is true
(all the possible outcomes)
Sample means
close to H_0:
high-probability values
if H_0 is true
Extreme, low-
probability values
if H_0 is true
μ from H_0
Extreme, low-
probability values
if H_0 is true


z
−1.96
0
1.96
−2.58
μ from H_0
2.58
−3.30
3.30
α = .05
α = .01
α = .001


Reject H_0
Reject H_0
Middle 95%:
High-probability values
if H_0 is true
80
μ from H_0
$z = -1.96$
0
$z = 1.96$
Critical region:

The independent samples t-test comes in two different forms:

The standard Student’s t-test, which assumes that the variance of the two groups are equal.

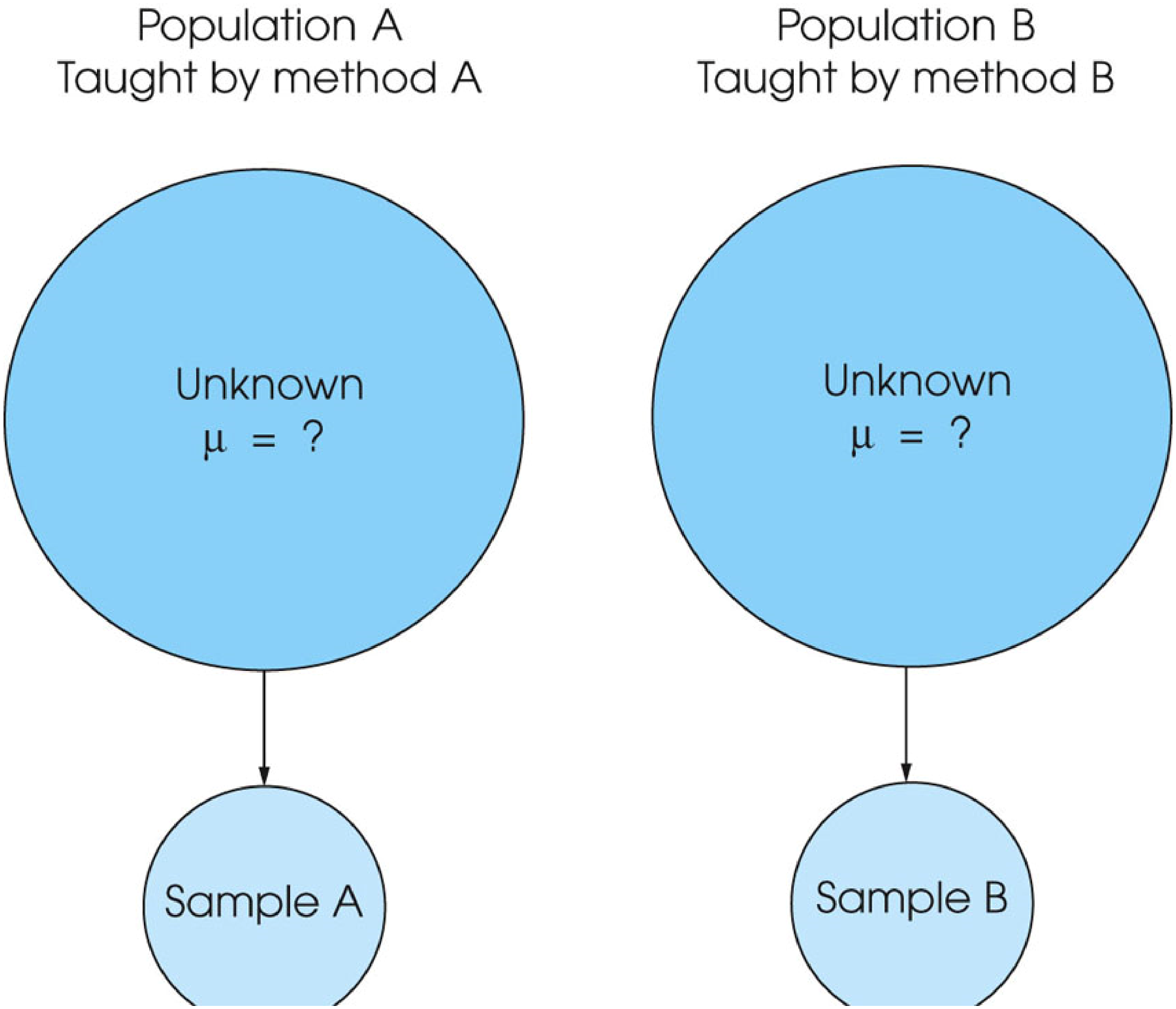
The Welch’s t-test, which is less restrictive compared to the original Student’s test. This is the test where you do not assume that the variance is the same in the two groups, which results in the fractional degrees of freedom.

The t-test is used in situations where a researcher has no prior knowledge about either of the two populations (or treatments) being compared.

In particular, the population means and standard deviations are all unknown.

Because the population variances are not known, these values must be estimated from the sample data.

If two samples are taken from the same population and are given exactly the same treatment, there still will be some difference between the sample means. This difference is called sampling error.


Population A
Taught by method A
Population B
Taught by method B
Unknown
μ = ?
Unknown
μ = ?
Sample A
Sample B

The general purpose of the t-test is to determine whether the sample mean difference obtained in a research study indicates a real mean difference between the two populations (or treatments) or whether the obtained difference is simply the result of sampling error.

The hypothesis test provides a standardized, formal procedure for determining whether the mean difference obtained in a research study is significantly greater than can be explained by sampling error.

The hypothesis test follows four-step procedure.

1. State the hypotheses and select an α level. For the t-test, H_0 states that there is no difference between the two population means.

2. Locate the critical region. The critical values for the t statistic are obtained using degrees of freedom that are determined by adding together the df value for the first sample and the df value for the second sample.

3. Compute the test statistic

Let $\mathcal{Q}_0$ and $\mathcal{Q}_1$ indicate two sets which are under the test.
Let also μ_0(resp. μ_1) and s_0^2 (resp. s_1^2) stand for the sample mean and sample variance of the set $\mathcal{Q}_0$ (resp. $\mathcal{Q}_1$), and n_0 and n_1 the cardinality of each set. The t-test statistic and the degree of freedom v are computed as

$$\boxed{\boldsymbol{t = \frac{\mu_0 - \mu_1}{\sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}}}}}$$

$$\boxed{\boldsymbol{v = \frac{\left(\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}\right)^2}{\frac{\left(\frac{s_0^2}{n_0}\right)^2}{n_0 - 1} + \frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1}}}}$$

In cases, where $s_0 \approx s_1$ and $n_0 \approx n_1$, the degree of freedom can be estimated by $v = n_0 + n_1 = n$

4. Make a decision. We estimate the probability to accept the null hypothesis by means of Student's t distribution density function,

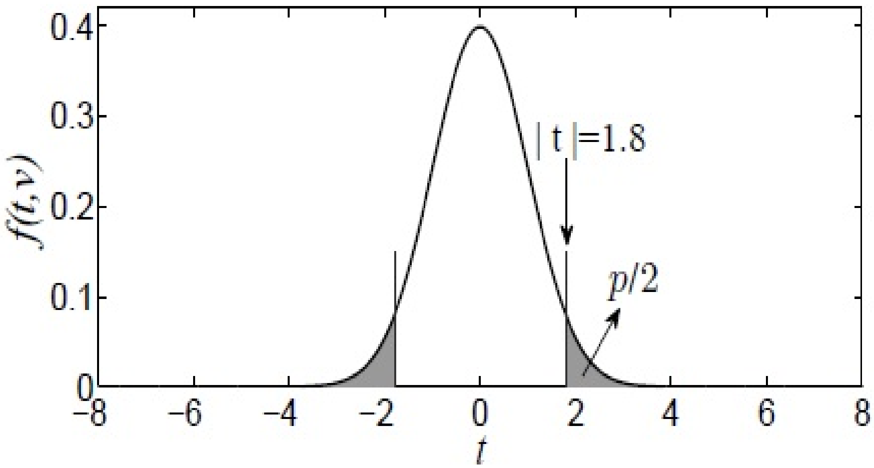
$$\boxed{\boldsymbol{f(t,v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}}}$$

Where $\boldsymbol{\Gamma(.)}$ denotes the gamma function and

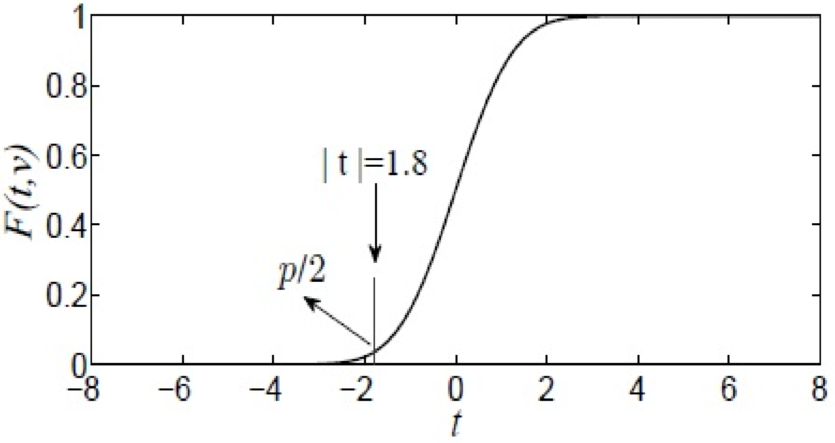
the desired probability is calculated as

$$\boxed{\boldsymbol{p = 2\int_{|t|}^{\infty} f(t,v)\,dt}}$$

probability density function



cumulative distribution function



small p values (alternatively big t values) give evidence to reject the null hypothesis and conclude that the sets were drawn from different populations.

For the sake of simplicity, usually a threshold $|t| > 4.5$ is defined to reject the null hypothesis without considering the degree of freedom and the aforementioned cumulative distribution function

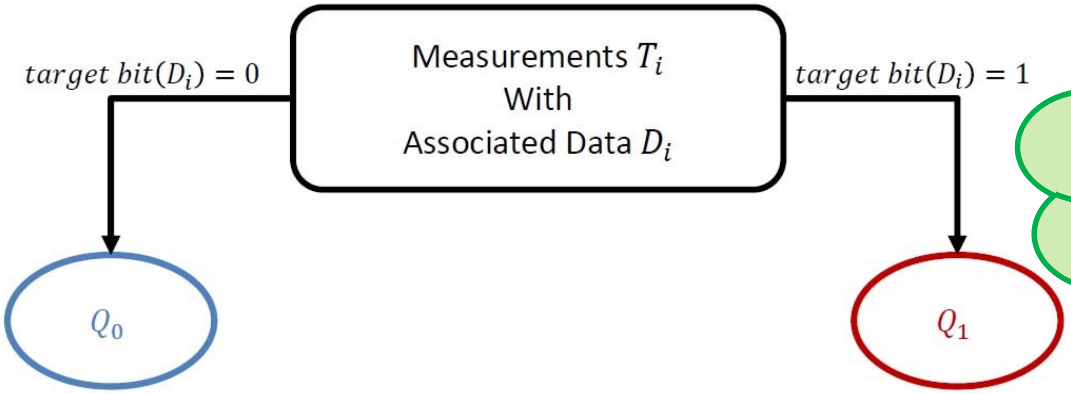
specific t-test

Consider n associated data (plain text or cipher text) $D_{i\in\{1,\cdots,n\}}$.

n side-channel measurements (traces $T_{i\in\{1,\cdots,n\}}$) are collected.

The device under test operates with a secret key that is kept constant.

Each trace $T_{i\in\{1,\cdots,n\}}$ containing m sample points $\{t_i^{(1)},\cdots,t_i^{(m)}\}$.



Measurements T_i
With
Associated Data D_i

$target\ bit(D_i) = 0$

$target\ bit(D_i) = 1$

Q_0

Q_1

If two sets are distinguishable from each other then DPA probable

The non-specific t-test examines the leakage of the DUT without performing an actual attack, and is in addition independent of its underlying architecture.

The test gives a level of confidence to conclude that the DUT has an exploitable leakage.

It indeed provides no information about the easiness/hardness of an attack which can exploit the leakage, nor about an appropriate intermediate value and the hypothetical model.

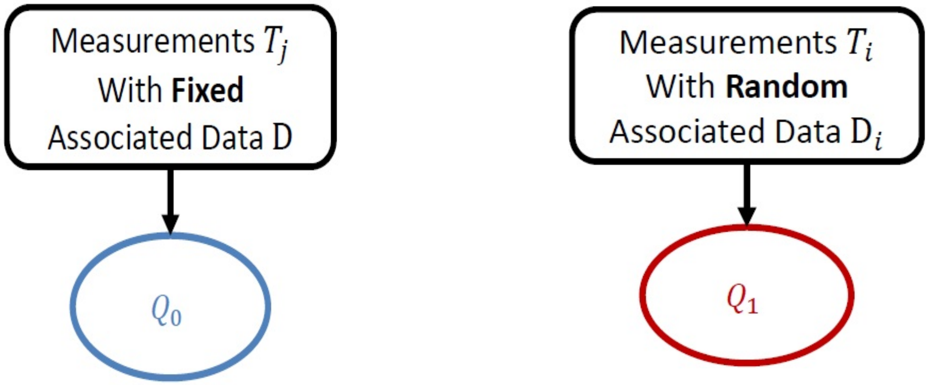
It can easily and rapidly report that the DUT fails to provide the desired security level, e.g., due to a mistake in the design engineering or a flaw in the countermeasure.

Non-specific t-test

A fixed associated data D is preselected.

A coin is filliped, and accordingly D **or** a fresh-randomly selected data is given to the DUT.

Side-channel measurements are collected.


Measurements T_j With **Fixed** Associated Data D
Q_0
Measurements T_i With **Random** Associated Data D_i
Q_1

The corresponding t-test is performed by categorizing the traces based on the associated data (D or random).

Such a test is also called fixed vs. random t-test.

If a non-specific t-test reports a detectable leakage, the specific one results in the same conclusion but with a higher confidence.

It may happen that a non-specific t-test by a certain D reports no exploitable leakage, but the same test using another D leads to the opposite conclusion.

Repeat a non-specific test with a couple of different D to avoid a false-positive conclusion.

The non-specific t-test can also be performed by a set of particular associated data ***D*** instead of a unique D. Such a non-specific t-test is also known as the semi-fixed vs. random test.

Question