Rijndael S-Box

Ji, Yong-hyeon

November 26, 2025

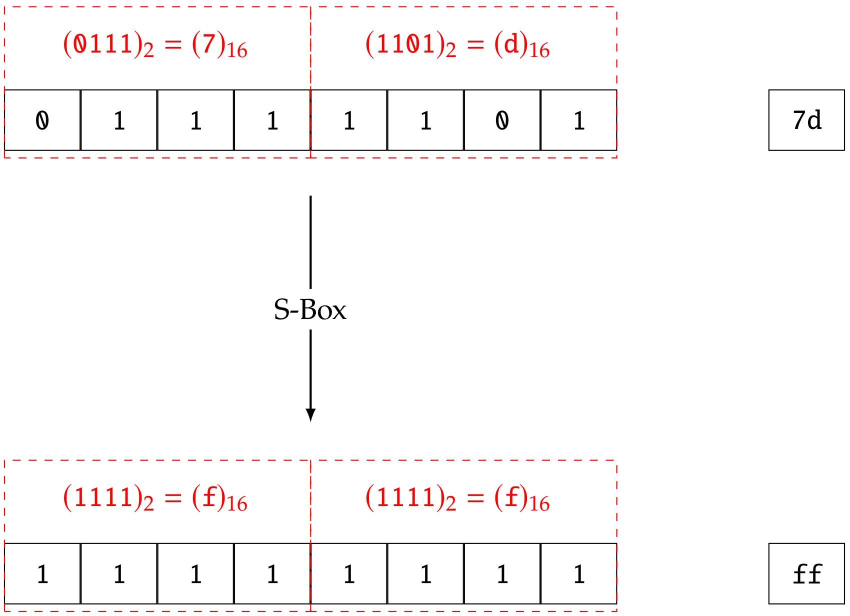
We cover the following topics in this note.

- Rijndael S-Box
- Vector Space $\mathbb{F}_2 = \{0, 1\}$

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
1	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
2	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
3	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
4	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
5	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
6	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
7	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
8	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
9	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
a	e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
b	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16

Table 1: AES S-box values in hexadecimal

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
1	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
2	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
3	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
4	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
5	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
6	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
7	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
8	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
9	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
a	e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
b	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16



We work in the finite field $\mathbb{F}_{2^8}$, represented as $\mathbb{F}_2^8$ via some fixed $\mathbb{F}_2$-linear isomorphism. The AES S-box (up to the special convention at 0) can be written as

$$S(x) = A(x^{-1}) + b,$$

where

- x^{-1} is the multiplicative inverse in $\mathbb{F}_{2^8}$, with the convention $0^{-1} := 0$,
    - $\mathbb{F}_{2^8} \cong \mathbb{F}_2[x]/\left(x^8 + x^4 + x^3 + x + 1\right)$
- $A : \mathbb{F}_2^8 \longrightarrow \mathbb{F}_2^8$ is an invertible $\mathbb{F}_2$-linear map (given by an 8×8 matrix over $\mathbb{F}_2$),
- $b \in \mathbb{F}_2^8$ is a fixed constant (an "offset").

$$\text{S-Box:} \qquad \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

This transformation is the sum of multiple rotations of the byte as a vector, where addition is the XOR operation:

$$s = b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4) \oplus \texttt{63}_{16}$$

where b represents the multiplicative inverse, $\oplus$ is the bitwise XOR operator, $\lll$ is a left bitwise circular shift, and the constant $\texttt{63}_{16} = \texttt{01100011}_2$ is given in hexadecimal.

b_7	b_6	b_5	b_4	b_3	b_2	b_1	b_0	$b = (b_7 b_6 \cdots b_0)_2$
b_6	b_5	b_4	b_3	b_2	b_1	b_0	b_7	$b \lll 1$
b_5	b_4	b_3	b_2	b_1	b_0	b_7	b_6	$b \lll 2$
b_4	b_3	b_2	b_1	b_0	b_7	b_6	b_5	$b \lll 3$
b_3	b_2	b_1	b_0	b_7	b_6	b_5	b_4	$b \lll 4$
s_7	s_6	s_5	s_4	s_3	s_2	s_1	s_0	$b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4)$

0.1 Hamming Weight and Hamming Distance on $\mathbb{F}_2^8$

Definition 1 (Hamming weight)**.** For a vector

$$x = (x_0, x_1, \ldots, x_7)^{\mathsf{T}} \in \mathbb{F}_2^8,$$

the *Hamming weight* of x is the number of nonzero coordinates:

$$\mathrm{wt}(x) := \bigl|\{\, i \in \{0, \ldots, 7\} : x_i = 1 \,\}\bigr|.$$

Definition 2 (Hamming distance)**.** For two vectors

$$x = (x_0, \ldots, x_7)^{\mathsf{T}}, \quad y = (y_0, \ldots, y_7)^{\mathsf{T}} \in \mathbb{F}_2^8,$$

the *Hamming distance* between x and y is

$$d_{\mathrm{H}}(x, y) := \bigl|\{\, i \in \{0, \ldots, 7\} : x_i \neq y_i \,\}\bigr|.$$

Since we are working over $\mathbb{F}_2$, we may equivalently write

$$d_{\mathrm{H}}(x, y) = \mathrm{wt}(x \oplus y),$$

where $x \oplus y$ denotes componentwise addition in $\mathbb{F}_2$ (bitwise XOR).

In particular, for the vector $b \in \mathbb{F}_2^8$ and its image Ab under the linear map with matrix A, the Hamming distance is

$$d_{\mathrm{H}}(b, Ab) = \mathrm{wt}\bigl(b \oplus Ab\bigr).$$

Likewise, for a cyclic left rotation $b \lll k$,

$$d_{\mathrm{H}}\bigl(b,\, b \lll k\bigr) = \mathrm{wt}\bigl(b \oplus (b \lll k)\bigr).$$

0.2 Concrete Examples of Hamming Distance

We work in $\mathbb{F}_2^8$.

Example 1 (Hamming distance between two arbitrary vectors)**.** Let

$$x = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad y = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{F}_2^8.$$

First compute their componentwise sum (XOR) in $\mathbb{F}_2$:

$$x \oplus y = \begin{bmatrix} 1 \oplus 0 \\ 0 \oplus 1 \\ 1 \oplus 1 \\ 1 \oplus 0 \\ 0 \oplus 0 \\ 0 \oplus 1 \\ 1 \oplus 0 \\ 0 \oplus 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

The Hamming weight of $x \oplus y$ is the number of 1's:

$$\mathrm{wt}(x \oplus y) = 5.$$

Therefore the Hamming distance between x and y is

$$d_{\mathrm{H}}(x, y) = \mathrm{wt}(x \oplus y) = 5.$$

Example 2 (Hamming distance between b and Ab)**.** Consider the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

and the vector

$$b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{F}_2^8.$$

Then

$$Ab = 1 \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

since b selects the first column of A.

Now compute

$$b \oplus Ab = \begin{bmatrix}1\\0\\0\\0\\0\\0\\0\\0\end{bmatrix} \oplus \begin{bmatrix}1\\1\\1\\1\\1\\0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\1\\1\\1\\1\\0\\0\\0\end{bmatrix}.$$

The Hamming weight is

$$\mathrm{wt}(b \oplus Ab) = 4,$$

hence the Hamming distance between b and Ab is

$$d_{\mathrm{H}}(b, Ab) = 4.$$

0.3 Why a Linear Transformation Acts as Diffusion

In the standard SPN (Substitution–Permutation Network) model, the state is a vector space over $\mathbb{F}_2$. For concreteness, let

$$V := \mathbb{F}_2^n$$

with the Hamming weight

$$\mathrm{wt}(x) := \big|\{i : x_i = 1\}\big|,$$

and the Hamming distance

$$d_{\mathrm{H}}(x, y) := \mathrm{wt}(x + y).$$

Definition 3 (Linear diffusion layer). A linear map $L : V \to V$, represented by an invertible matrix $M \in \mathrm{GL}(n, 2)$, is called a *diffusion layer* if it spreads nonzero inputs so that $\mathrm{wt}(L(x))$ is typically large whenever $x \neq 0$.

Intuitively, a change in a small number of input bits should affect many output bits.

1. Componentwise view: each output bit mixes many input bits

Let $L(x) = Mx$ with $M = (M_{ij})$. The i-th output bit is

$$(L(x))_i = \sum_{j=1}^{n} M_{ij} x_j \in \mathbb{F}_2.$$

Thus, row i of M determines which input bits influence output bit i:

$$(L(x))_i = \bigoplus_{j : M_{ij}=1} x_j.$$

If each row of M has many 1's, then each output bit is the XOR of many input bits. In particular, flipping a single input bit x_k changes all output bits i such that $M_{ik} = 1$. Therefore, the number of output bits affected by a single-bit change in position k is

$$\mathrm{wt}\big(Me_k\big),$$

where e_k is the k-th standard basis vector. A good diffusion layer satisfies

$$\mathrm{wt}(Me_k) \text{ is large for all } k.$$

2. Hamming weight and branch number

A deeper global measure of diffusion is the *branch number*.

Definition 4 (Branch number)**.** Let $L(x) = Mx$ be linear on $\mathbb{F}_2^n$. The branch number of L is

$$B(L) := \min_{x \in V \setminus \{0\}} \big(\mathrm{wt}(x) + \mathrm{wt}(L(x))\big).$$

If $B(L)$ is large, then for any nonzero input difference x we cannot have both $\mathrm{wt}(x)$ and $\mathrm{wt}(L(x))$ small at the same time. In other words:

$$x \neq 0 \implies \mathrm{wt}(x) + \mathrm{wt}(L(x)) \geq B(L).$$

Cryptographically, this implies:

- If the input difference is concentrated in a few bits (small $\mathrm{wt}(x)$), then the output difference must be spread over many bits (large $\mathrm{wt}(L(x))$).
- In an SPN with parallel S-boxes, L acts between nonlinear layers. A large branch number forces many S-boxes to become "active" in the following rounds, which raises the cost of any differential or linear trail.

Thus, the branch number is a *purely linear-algebraic* quantity that quantifies how well the matrix M performs diffusion.

3. Linearity and propagation of differences

Let $L(x) = Mx$ be linear and consider two inputs $x, y \in V$. The difference of outputs is

$$L(x) + L(y) = L(x + y).$$

Hence, the evolution of *differences* under L is exactly the evolution of single vectors under the same linear map. In particular, if $\Delta = x + y$, then

$$\Delta_{\mathrm{out}} = L(x) + L(y) = L(\Delta).$$

Therefore, the way small input differences spread through the cipher is controlled entirely by the linear map L between the nonlinear S-box layers. Good diffusion means: for all nonzero Δ, the Hamming weight $\mathrm{wt}(L(\Delta))$ is large, and the branch number $B(L)$ is high.

4. Why linear?

There are several reasons why diffusion is implemented *linearly*:

- **Simplicity and efficiency.** Matrix multiplication over $\mathbb{F}_2$ is easy to implement in both hardware and software (XORs and bit shifts).

- **Invariance of nonlinear strength.** Composing a nonlinear S-box with an invertible linear map L (before or after) yields an *affine-equivalent* S-box with the same nonlinearity and differential uniformity. Thus one can design nonlinearity (via inversion in $\mathbb{F}_{2^8}$, for instance) and diffusion (via a carefully chosen M) largely independently.

- **Clean analysis.** Because L is linear, its effect on differences (and on linear masks) is completely determined by standard linear algebra over $\mathbb{F}_2$. This makes it possible to prove lower bounds on the number of active S-boxes in multiple rounds (the "wide trail" strategy).

5. Example: the given 8×8 matrix

For the specific matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

we have:

- $A \in \mathrm{GL}(8,2)$: it is invertible over $\mathbb{F}_2$, so no information is lost.

- Each row has 5 ones: every output bit is the XOR of 5 input bits.

- The branch number is
$$B(A) = \min_{x \neq 0} \big(\mathrm{wt}(x) + \mathrm{wt}(Ax)\big) = 4,$$
so no nonzero difference can remain concentrated in too few bits in both input and output simultaneously.

These are precisely the linear-algebraic features that make a matrix like A serve as a diffusion layer in a cipher: it spreads any local change across many coordinates and, when iterated between nonlinear layers, forces a large part of the state to be affected.

0.4 A Geometric Viewpoint in Finite Dimension

We work over the vector space

$$V := \mathbb{F}_2^8.$$

The elements of V may be viewed as the vertices of the 8-dimensional Hamming cube. Concretely,

$$V = \{0,1\}^8,$$

and we equip V with the Hamming distance

$$d(x,y) := \big|\{i \in \{0,\dots,7\} : x_i \neq y_i\}\big|.$$

Two vertices are joined by an edge if and only if they differ in exactly one coordinate (i.e. distance 1).

The linear map associated to the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

is the map $T : V \to V$ defined by

$$T(b) = Ab.$$

Equivalently, in terms of cyclic left rotations (indices modulo 8),

$$T(b) = b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4),$$

so that each coordinate satisfies

$$(T(b))_i = b_i \oplus b_{i-1} \oplus b_{i-2} \oplus b_{i-3} \oplus b_{i-4} \quad (i \in \mathbb{Z}_8).$$

- As a linear transformation, T is a bijection of V, i.e. $A \in \mathrm{GL}(8,2)$. In geometric terms, T is an *automorphism* of the finite affine space $\mathrm{AG}(8,2)$.
- Lines in V are affine 1-dimensional subspaces of the form $\ell = x + \mathbb{F}_2 v$, where $v \neq 0$. Since T

is linear, it maps lines to lines:

$$T(\ell) = T(x) + \mathbb{F}_2\,T(v).$$

More generally, every affine subspace (finite "flat") is mapped to an affine subspace of the same dimension.

- Thus T preserves the entire incidence structure of the Hamming cube: it permutes vertices, carries edges to edges, and maps affine k-dimensional subspaces to affine k-dimensional subspaces. In this sense, T is a geometric symmetry of the finite space $\mathbb{F}_2^8$.

0.5 Why the Expression $b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4)$ Corresponds to the Matrix A

We work over the vector space $\mathbb{F}_2^8$. Let

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} \in \mathbb{F}_2^8,$$

and interpret the index set $\{0, 1, \ldots, 7\}$ as the cyclic group $\mathbb{Z}_8$. Addition of indices is always taken modulo 8.

Definition 5 (Cyclic left rotation)**.** For $k \in \mathbb{Z}_8$, the cyclic left rotation of b by k positions, denoted $b \lll k$, is the vector $b' \in \mathbb{F}_2^8$ with

$$(b \lll k)_i := b_{\,i-k}, \qquad i \in \mathbb{Z}_8,$$

where the index $i - k$ is taken modulo 8.

Thus, componentwise, we have

$$(b \lll 0)_i = b_i, \quad (b \lll 1)_i = b_{i-1}, \quad (b \lll 2)_i = b_{i-2}, \quad (b \lll 3)_i = b_{i-3}, \quad (b \lll 4)_i = b_{i-4}.$$

Consider the linear map

$$T : \mathbb{F}_2^8 \to \mathbb{F}_2^8, \qquad T(b) := b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4),$$

where $\oplus$ denotes addition in $\mathbb{F}_2$ (bitwise XOR).

Step 1: Coordinate formula for $T(b)$. Let $c := T(b)$, so $c = (c_0, \ldots, c_7)^\mathsf{T}$. Then for each $i \in \mathbb{Z}_8$,

$$\begin{aligned} c_i &= (b)_i \oplus (b \lll 1)_i \oplus (b \lll 2)_i \oplus (b \lll 3)_i \oplus (b \lll 4)_i \\ &= b_i \oplus b_{i-1} \oplus b_{i-2} \oplus b_{i-3} \oplus b_{i-4}. \end{aligned}$$

In other words,

$$c_i = \sum_{k=0}^{4} b_{\,i-k} \quad \text{in } \mathbb{F}_2.$$

Explicitly, writing out the eight coordinates:

$$\begin{aligned} c_0 &= b_0 \oplus b_7 \oplus b_6 \oplus b_5 \oplus b_4, \\ c_1 &= b_1 \oplus b_0 \oplus b_7 \oplus b_6 \oplus b_5, \\ c_2 &= b_2 \oplus b_1 \oplus b_0 \oplus b_7 \oplus b_6, \\ c_3 &= b_3 \oplus b_2 \oplus b_1 \oplus b_0 \oplus b_7, \\ c_4 &= b_4 \oplus b_3 \oplus b_2 \oplus b_1 \oplus b_0, \\ c_5 &= b_5 \oplus b_4 \oplus b_3 \oplus b_2 \oplus b_1, \\ c_6 &= b_6 \oplus b_5 \oplus b_4 \oplus b_3 \oplus b_2, \\ c_7 &= b_7 \oplus b_6 \oplus b_5 \oplus b_4 \oplus b_3. \end{aligned}$$

Step 2: Read off the matrix entries. Let A denote the matrix of the linear map T with respect to the standard basis of $\mathbb{F}_2^8$. By definition, the entries $A_{i,j}$ satisfy

$$c_i = (T(b))_i = \sum_{j=0}^{7} A_{i,j} b_j,$$

with addition in $\mathbb{F}_2$.

From the coordinate formula

$$c_i = b_i \oplus b_{i-1} \oplus b_{i-2} \oplus b_{i-3} \oplus b_{i-4},$$

we see that

$$A_{i,j} = \begin{cases} 1, & \text{if } j \in \{i,\ i-1,\ i-2,\ i-3,\ i-4\} \bmod 8, \\ 0, & \text{otherwise.} \end{cases}$$

Writing out the rows explicitly (with indices $0, \ldots, 7$ in order), we obtain

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Conclusion. By construction, for every $b \in \mathbb{F}_2^8$,

$$T(b) = b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4)$$

and

$$T(b) = Ab.$$

Therefore, the expression

$$b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4)$$

is exactly the action of the matrix A on the vector b in the standard basis of $\mathbb{F}_2^8$.

$$A\begin{bmatrix} b_0\\ b_1\\ b_2\\ b_3\\ b_4\\ b_5\\ b_6\\ b_7 \end{bmatrix} = b_0\begin{bmatrix}1\\1\\1\\1\\1\\0\\0\\0\end{bmatrix} + b_1\begin{bmatrix}0\\1\\1\\1\\1\\1\\0\\0\end{bmatrix} + b_2\begin{bmatrix}0\\0\\1\\1\\1\\1\\1\\0\end{bmatrix} + b_3\begin{bmatrix}0\\0\\0\\1\\1\\1\\1\\1\end{bmatrix} + b_4\begin{bmatrix}1\\0\\0\\0\\1\\1\\1\\1\end{bmatrix} + b_5\begin{bmatrix}1\\1\\0\\0\\0\\1\\1\\1\end{bmatrix} + b_6\begin{bmatrix}1\\1\\1\\0\\0\\0\\1\\1\end{bmatrix} + b_7\begin{bmatrix}1\\1\\1\\1\\0\\0\\0\\1\end{bmatrix}.$$

We write

$$b = \begin{bmatrix} b_0\\ b_1\\ b_2\\ b_3\\ b_4\\ b_5\\ b_6\\ b_7 \end{bmatrix} \in \mathbb{F}_2^8.$$

Then each coordinate of Ab is a linear combination of $\{b_0, \dots, b_7\}$ with coefficients in $\mathbb{F}_2$:

$$Ab = \begin{bmatrix} c_0\\ c_1\\ c_2\\ c_3\\ c_4\\ c_5\\ c_6\\ c_7 \end{bmatrix} = \begin{bmatrix} b_0 + b_4 + b_5 + b_6 + b_7\\ b_0 + b_1 + b_5 + b_6 + b_7\\ b_0 + b_1 + b_2 + b_6 + b_7\\ b_0 + b_1 + b_2 + b_3 + b_7\\ b_0 + b_1 + b_2 + b_3 + b_4\\ b_1 + b_2 + b_3 + b_4 + b_5\\ b_2 + b_3 + b_4 + b_5 + b_6\\ b_3 + b_4 + b_5 + b_6 + b_7 \end{bmatrix}.$$

Equivalently, in compact index notation (indices taken modulo 8),

$$(Ab)_i = \sum_{k=0}^{4} b_{\,i-k} \quad \text{for } i = 0, 1, \dots, 7,$$

so every entry $(Ab)_i$ is a linear combination of the scalars $b_0, \dots, b_7$ with coefficients in $\{0, 1\} \subset \mathbb{F}_2$.

0.6 Bitwise Description via Cyclic Left Rotations

Let

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} \in \mathbb{F}_2^8.$$

We interpret b as an 8-bit word, where b_i is the bit in position i, with indices taken in $\mathbb{Z}_8 = \{0, 1, \dots, 7\}$.

Definition 6 (Cyclic left rotation). For $k \in \mathbb{Z}_8$, the *cyclic left rotation* of b by k bits, denoted $b \lll k$, is the vector in $\mathbb{F}_2^8$ whose i-th component is

$$(b \lll k)_i := b_{\,i-k \bmod 8}, \qquad i \in \mathbb{Z}_8.$$

With this convention, the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

acts on b by

$$Ab = \begin{bmatrix} b_0 + b_4 + b_5 + b_6 + b_7 \\ b_0 + b_1 + b_5 + b_6 + b_7 \\ b_0 + b_1 + b_2 + b_6 + b_7 \\ b_0 + b_1 + b_2 + b_3 + b_7 \\ b_0 + b_1 + b_2 + b_3 + b_4 \\ b_1 + b_2 + b_3 + b_4 + b_5 \\ b_2 + b_3 + b_4 + b_5 + b_6 \\ b_3 + b_4 + b_5 + b_6 + b_7 \end{bmatrix},$$

where addition is in $\mathbb{F}_2$ (bitwise XOR).

Equivalently, each component satisfies

$$(Ab)_i = b_i + b_{i-1} + b_{i-2} + b_{i-3} + b_{i-4} \quad \text{for all } i \in \mathbb{Z}_8,$$

so that

$$Ab = b \oplus (b \lll 1) \oplus (b \lll 2) \oplus (b \lll 3) \oplus (b \lll 4),$$

where $\oplus$ denotes bitwise XOR and $\lll k$ is 8-bit cyclic left rotation by k positions.

1 Computing Ab in Several Ways

We consider the matrix

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

and the column vector

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix}.$$

We write

$$Ab = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{bmatrix}.$$

1.1 Direct Row-by-Row Computation

Each component c_i is the dot product of the i-th row of A with the vector b. Thus

$$\begin{aligned}
c_0 &= b_0 + b_4 + b_5 + b_6 + b_7,\\
c_1 &= b_0 + b_1 + b_5 + b_6 + b_7,\\
c_2 &= b_0 + b_1 + b_2 + b_6 + b_7,\\
c_3 &= b_0 + b_1 + b_2 + b_3 + b_7,\\
c_4 &= b_0 + b_1 + b_2 + b_3 + b_4,\\
c_5 &= b_1 + b_2 + b_3 + b_4 + b_5,\\
c_6 &= b_2 + b_3 + b_4 + b_5 + b_6,\\
c_7 &= b_3 + b_4 + b_5 + b_6 + b_7.
\end{aligned}$$

Here the addition can be understood either as ordinary addition (over $\mathbb{R}$) or as addition in $\mathbb{F}_2$ (bitwise XOR).

In vector form,

$$Ab = \begin{bmatrix} b_0 + b_4 + b_5 + b_6 + b_7\\ b_0 + b_1 + b_5 + b_6 + b_7\\ b_0 + b_1 + b_2 + b_6 + b_7\\ b_0 + b_1 + b_2 + b_3 + b_7\\ b_0 + b_1 + b_2 + b_3 + b_4\\ b_1 + b_2 + b_3 + b_4 + b_5\\ b_2 + b_3 + b_4 + b_5 + b_6\\ b_3 + b_4 + b_5 + b_6 + b_7 \end{bmatrix}.$$

1.2 Interpretation over $\mathbb{F}_2$ (XOR Form)

If $b_i \in \{0,1\}$ and we work over the field $\mathbb{F}_2$, we usually write $\oplus$ for addition (XOR). Then

$$\begin{aligned}
c_0 &= b_0 \oplus b_4 \oplus b_5 \oplus b_6 \oplus b_7,\\
c_1 &= b_0 \oplus b_1 \oplus b_5 \oplus b_6 \oplus b_7,\\
c_2 &= b_0 \oplus b_1 \oplus b_2 \oplus b_6 \oplus b_7,\\
c_3 &= b_0 \oplus b_1 \oplus b_2 \oplus b_3 \oplus b_7,\\
c_4 &= b_0 \oplus b_1 \oplus b_2 \oplus b_3 \oplus b_4,\\
c_5 &= b_1 \oplus b_2 \oplus b_3 \oplus b_4 \oplus b_5,\\
c_6 &= b_2 \oplus b_3 \oplus b_4 \oplus b_5 \oplus b_6,\\
c_7 &= b_3 \oplus b_4 \oplus b_5 \oplus b_6 \oplus b_7.
\end{aligned}$$

1.3 Cyclic Convolution Description

Let the indices be taken modulo 8, i.e. we regard $\{0, 1, \ldots, 7\}$ as $\mathbb{Z}_8$. Then each output component c_i is a sum of five consecutive inputs:

$$c_i = \sum_{k=0}^{4} b_{\,i-k \bmod 8}, \qquad i \in \{0, 1, \ldots, 7\}.$$

In other words, the linear map $b \mapsto Ab$ is the *circular convolution* of the sequence

$$(b_0, b_1, \ldots, b_7)$$

with the *kernel*

$$h = (1, 1, 1, 1, 1, 0, 0, 0).$$

1.4 Polynomial Representation in $\mathbb{F}_2[x]/(x^8 - 1)$

Encode the vector b as a polynomial

$$B(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_7 x^7$$

with coefficients in $\mathbb{F}_2$. Define

$$H(x) = 1 + x + x^2 + x^3 + x^4.$$

Then the product Ab corresponds to the polynomial

$$C(x) = H(x)\, B(x) \bmod (x^8 - 1),$$

and the coefficients of $C(x)$ are precisely the components $c_0, \ldots, c_7$ of the vector Ab.

Thus the single matrix–vector product

$$Ab$$

admits equivalent descriptions as:

- an explicit linear combination of the b_i,
- a bitwise XOR of selected coordinates (over $\mathbb{F}_2$),
- a cyclic convolution with the kernel $(1, 1, 1, 1, 1, 0, 0, 0)$,
- multiplication by $1 + x + x^2 + x^3 + x^4$ in the quotient ring $\mathbb{F}_2[x]/(x^8 - 1)$.

2 The Ambient Vector Space

We work throughout over the finite field

$$\mathbb{F}_2 = \{0, 1\}$$

with addition and multiplication taken modulo 2.

Definition 7. Let

$$V := \mathbb{F}_2^8$$

denote the 8-dimensional vector space of column vectors with entries in $\mathbb{F}_2$. An element

$$x \in V$$

is a column vector

$$x = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_7 \end{pmatrix}, \quad x_i \in \mathbb{F}_2,$$

with addition and scalar multiplication defined componentwise over $\mathbb{F}_2$.

Definition 8. The *standard basis* of V is the family

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad i = 0, \dots, 7,$$

where the entry 1 is in position i and all other entries are 0.

Every $x \in V$ has a unique representation

$$x = \sum_{i=0}^{7} x_i e_i.$$

3 Linear Maps and Invertibility

Definition 9. A map $L : V \to V$ is called *linear* if for all $x, y \in V$ and all $\alpha, \beta \in \mathbb{F}_2$,

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y).$$

With respect to the standard basis of V, every linear map L is represented by a unique matrix

$$M \in \mathrm{Mat}_{8\times 8}(\mathbb{F}_2)$$

such that

$$L(x) = Mx$$

for all $x \in V$.

Definition 10. A linear map $L : V \to V$ is called *invertible* if there exists a linear map $L^{-1} : V \to V$ such that

$$L^{-1} \circ L = \mathrm{id}_V \quad \text{and} \quad L \circ L^{-1} = \mathrm{id}_V.$$

Lemma 1. *Let $M \in \mathrm{Mat}_{8\times 8}(\mathbb{F}_2)$. The following are equivalent:*

1. *The map $x \mapsto Mx$ is invertible as a linear map $V \to V$.*
2. $\ker(M) = \{0\}$.
3. $\mathrm{rank}(M) = 8$.
4. $\det(M) = 1 \in \mathbb{F}_2$.

The group of invertible 8×8 matrices over $\mathbb{F}_2$ is denoted

$$\mathrm{GL}(8, 2) := \{M \in \mathrm{Mat}_{8\times 8}(\mathbb{F}_2) : \det(M) = 1\}.$$

4 Affine Transformations of V

Definition 11. A map $A : V \to V$ is called *affine* if there exists a linear map $L : V \to V$ and a vector $c \in V$ such that

$$A(x) = L(x) + c \quad \text{for all } x \in V.$$

The linear map L is called the *linear part* of A.

In coordinates, if $L(x) = Mx$ for some $M \in \mathrm{Mat}_{8\times 8}(\mathbb{F}_2)$, then

$$A(x) = Mx + c, \quad c \in V.$$

Proposition 2. *Let $A : V \to V$ be given by*

$$A(x) = Mx + c$$

with $M \in \mathrm{Mat}_{8\times 8}(\mathbb{F}_2)$ and $c \in V$. Then

$$A \text{ is bijective} \iff M \in \mathrm{GL}(8, 2).$$

Proof. Suppose first that $M \in \mathrm{GL}(8,2)$. Then we define a map $A^{-1} : V \to V$ by

$$A^{-1}(y) := M^{-1}(y - c).$$

This is well-defined and affine. For all $x \in V$,

$$A^{-1}(A(x)) = M^{-1}(Mx + c - c) = M^{-1}Mx = x,$$

and for all $y \in V$,

$$A(A^{-1}(y)) = M\big(M^{-1}(y - c)\big) + c = (y - c) + c = y.$$

Thus A is bijective.

Conversely, suppose A is bijective. Let $x, x' \in V$ satisfy $Mx = Mx'$. Then

$$A(x) - c = Mx = Mx' = A(x') - c,$$

so $A(x) = A(x')$. By injectivity of A, it follows that $x = x'$. Hence the linear map $x \mapsto Mx$ is injective and thus invertible. Therefore $M \in \mathrm{GL}(8,2)$. □

5 The Rijndael S-Box as a Composition

5.1 The Nonlinear Core

The Rijndael (AES) S-box is a permutation

$$S : V \to V$$

constructed as the composition of two conceptually distinct steps:

1. a nonlinear permutation $N : V \to V$, and
2. an affine transformation $A_{\mathrm{aff}}(x) = Mx + c$, with $M \in \mathrm{GL}(8,2)$, $c \in V$.

To define N, one first identifies V with the finite field $\mathbb{F}_{2^8}$ via a fixed $\mathbb{F}_2$-vector-space isomorphism

$$\varphi : \mathbb{F}_{2^8} \to V.$$

Definition 12. Define a permutation $N : V \to V$ by

$$N(x) = \begin{cases} \varphi\big((\varphi^{-1}(x))^{-1}\big), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

The map N is the multiplicative inverse on $\mathbb{F}_{2^8}$, expressed in the coordinate representation V. It is not linear over $\mathbb{F}_2$.

5.2 Affine Output Transformation

The Rijndael S-box is then defined by

$$S(x) := M \cdot N(x) + c,$$

where $M \in \mathrm{GL}(8,2)$ and $c \in V$ are fixed. By Proposition 2 and the fact that N is a permutation, we see that S is a permutation:

$$S = A_{\text{aff}} \circ N,$$

and the composition of two bijections is a bijection.

From a linear algebraic perspective, once the nonlinear permutation N is fixed, the S-box is determined entirely by the choice of the pair (M, c), with $M \in \mathrm{GL}(8,2)$.

6 Linear-Algebraic Notions of Diffusion

6.1 Hamming Weight and Bit Diffusion

Definition 13. For $x = (x_0, \dots, x_7)^\top \in V$, the *Hamming weight* of x is the integer

$$\mathrm{wt}(x) := \left|\{i \in \{0, \dots, 7\} : x_i = 1\}\right|.$$

Definition 14. Let $L : V \to V$ be linear, $L(x) = Mx$. We say that L has good *bit-wise diffusion* if:

1. For every nonzero standard basis vector e_i, the weight $\mathrm{wt}(L(e_i))$ is relatively large.
2. For every nonzero $x \in V$, the weight $\mathrm{wt}(L(x))$ is bounded below by a prescribed constant.

The above conditions depend only on the matrix M and are purely linear in nature.

6.2 Difference Propagation Through the Affine Layer

Let $S(x) = MN(x) + c$ be the S-box. For any $\Delta x \in V$ and all $x \in V$,

$$\begin{aligned} S(x + \Delta x) + S(x) &= MN(x + \Delta x) + c + MN(x) + c \\ &= M\big(N(x + \Delta x) + N(x)\big). \end{aligned}$$

Thus the difference

$$\Delta S := S(x + \Delta x) + S(x)$$

is obtained by applying the linear map M to the *nonlinear difference*

$$\Delta N := N(x+\Delta x) + N(x).$$

Lemma 3. *For all $x, \Delta x \in V$,*

$$S(x+\Delta x) + S(x) = M\big(N(x+\Delta x) + N(x)\big).$$

In particular, the mapping $\Delta N \mapsto \Delta S$ is entirely determined by the linear map M.

6.3 Branch Number

Definition 15. Let $L : V \to V$ be linear, $L(x) = Mx$. The *branch number* of L is defined as

$$B(L) := \min_{x \in V \setminus \{0\}} \big(\mathrm{wt}(x) + \mathrm{wt}(L(x))\big).$$

A large branch number means that for every nonzero x, it is impossible for both x and $L(x)$ to have simultaneously small Hamming weight. This is a purely linear property of the matrix M.

In the context of the S-box, once the nonlinear part N is fixed, the branch number of the affine layer $L(x) = Mx$ is a fundamental linear-algebraic measure of how effectively bit patterns are spread.

7 Affine Transformations and Change of Basis

7.1 Change of Basis in V

Let $\mathcal{B} = \{b_0, \ldots, b_7\}$ be a basis of V. Every $x \in V$ can be written uniquely as

$$x = \sum_{i=0}^{7} \alpha_i b_i, \quad \alpha_i \in \mathbb{F}_2.$$

We define the coordinate column vector of x with respect to $\mathcal{B}$ as

$$[x]_{\mathcal{B}} := \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_7 \end{pmatrix} \in V.$$

If $\mathcal{B}' = \{b'_0, \ldots, b'_7\}$ is another basis, then there exists a unique matrix $P \in \mathrm{GL}(8,2)$ such that

$$[x]_{\mathcal{B}'} = P\,[x]_{\mathcal{B}}$$

for all $x \in V$. Thus every change of basis in V is represented by left-multiplication by an element of $\mathrm{GL}(8,2)$.

7.2 Conjugation of the Nonlinear Core

Let $N : V \to V$ be the nonlinear permutation induced by inversion in $\mathbb{F}_{2^8}$, as defined earlier. If we change coordinates by a matrix $P \in \mathrm{GL}(8,2)$, then the coordinate representation of N in the new basis is the conjugate map

$$N'(x) := P\,N(P^{-1}x).$$

Thus composing N with invertible linear maps on the left and right corresponds to viewing the same abstract permutation under different coordinate systems.

The affine post-transformation used in the Rijndael S-box,

$$A_{\mathrm{aff}}(x) = Mx + c,$$

is therefore not merely an implementational detail: its linear part M selects a particular way of mixing the coordinates of the output of the nonlinear core N before presenting them as the final S-box output bits.

8 Linear-Algebraic Constraints on the Rijndael Matrix

From a purely linear algebraic point of view, the choice of the Rijndael S-box matrix M can be described as the selection of an element of $\mathrm{GL}(8,2)$ subject to certain constraints. Typical such constraints include:

1. **Invertibility.** One requires
$$M \in \mathrm{GL}(8,2),$$
i.e. $\det(M) = 1$ in $\mathbb{F}_2$. By Proposition 2, this ensures that the affine map $x \mapsto Mx + c$ is bijective, hence preserves the permutation property of the S-box.

2. **Non-degenerate bit influence.** For each standard basis vector e_i,
$$\mathrm{wt}(Me_i)$$
should be relatively large, so that changing a single input bit (after the nonlinear map N) influences many of the output bits of the affine layer.

3. **Absence of trivial linear relations.** One typically requires that certain subsets of rows or columns of M are linearly independent, or that submatrices of M have sufficiently large rank, in order to avoid simple linear relations among output bits in terms of a small number of input bits.

4. **Good branch number.** One seeks to maximize

$$B(M) := \min_{x \in V \setminus \{0\}} \big(\mathrm{wt}(x) + \mathrm{wt}(Mx)\big),$$

subject to the above constraints and to implementational considerations. This ensures that any nontrivial input pattern to the affine layer yields an output pattern whose sparsity cannot be simultaneously small.

All of these constraints are expressed purely in terms of linear algebra over $\mathbb{F}_2$ and are independent of the cryptanalytic details of the nonlinear part N.

9 The Importance of the 8×8 Matrix

From a linear algebraic perspective, the Rijndael S-box

$$S(x) = MN(x) + c$$

is determined, once the nonlinear permutation N is fixed, by the choice of a single pair (M, c) with $M \in \mathrm{GL}(8, 2)$, $c \in V$. Among these, the constant vector c serves to translate all outputs uniformly, while the matrix M governs the structural properties of the S-box that are visible at the level of linear algebra. Concretely:

- *Affine structure.* Every affine map $A_{\mathrm{aff}} : V \to V$ has the form $A_{\mathrm{aff}}(x) = Mx+c$. Up to translation, the entire affine structure is captured by the single matrix M.
- *Bijectivity.* The condition $M \in \mathrm{GL}(8, 2)$, equivalently $\det(M) = 1$, is necessary and sufficient for the S-box to remain a permutation, since N is bijective.
- *Diffusion.* For any difference $\Delta x \in V$,

  $$S(x + \Delta x) + S(x) = M\big(N(x + \Delta x) + N(x)\big),$$

  so the propagation of differences through the affine layer—as measured, for example, by Hamming weight or branch number—is entirely governed by the linear map M.
- *Coordinate system.* Since $\mathrm{GL}(8, 2)$ acts as the group of changes of basis on V, the choice of M specifies a particular coordinate system in which the outputs of N are expressed. In this sense, M encodes a distinguished way of mixing and presenting the eight coordinate bits after the nonlinear step.

In summary, from a purely linear algebraic standpoint, the Rijndael S-box is the composition of a fixed nonlinear permutation N with an affine transformation whose linear part is a carefully chosen element $M \in \mathrm{GL}(8,2)$. All of the structural, coordinate-based properties of the S-box that can be articulated in linear algebraic terms—such as invertibility, diffusion, absence of simple linear relations, and behavior of Hamming weights—are encoded in this single 8×8 matrix.

10 A Structured 8×8 Binary Matrix

We consider the 8×8 matrix

$$M := \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

All entries of M lie in $\{0, 1\}$. Thus M may be regarded:

- either as a matrix with real entries, $M \in \mathrm{Mat}_{8\times 8}(\mathbb{R})$,
- or as a matrix with entries in the finite field $\mathbb{F}_2$, $M \in \mathrm{Mat}_{8\times 8}(\mathbb{F}_2)$.

In both cases, M defines a linear operator on an 8-dimensional vector space by left multiplication.

10.1 As a Linear Operator on $\mathbb{K}^8$

Let $\mathbb{K}$ be a field, and consider the vector space

$$V := \mathbb{K}^8$$

of column vectors of length 8 over $\mathbb{K}$. We equip V with its standard basis

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \ldots, \ e_8 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Definition 16. Define the linear operator

$$T : V \to V, \qquad T(x) := Mx.$$

With respect to the standard basis $\{e_1, \dots, e_8\}$, the matrix of T is precisely M. The j-th column of M is the coordinate vector of $T(e_j)$:

$$T(e_j) = Me_j = \text{the } j\text{-th column of } M, \quad j = 1, \dots, 8.$$

10.2 Row and Column Structure

Each row of M contains exactly three entries equal to 1 and five entries equal to 0. Concretely, if we denote by $\mathbf{1}$ the all-ones column vector

$$\mathbf{1} := \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \in V,$$

then a direct computation shows that over $\mathbb{R}$

$$M\mathbf{1} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} = 3\,\mathbf{1},$$

and for every row i,

$$\sum_{j=1}^{8} M_{ij} = 3.$$

Similarly, each column of M has exactly three entries equal to 1, so

$$\sum_{i=1}^{8} M_{ij} = 3 \quad \text{for all } j = 1, \dots, 8.$$

In particular, the in-degree and out-degree (see the graph interpretation below) are both equal to 3 for every index.

10.3 Invertibility and Rank

We now record basic invariants of M.

Proposition 4. *Viewed as a matrix over $\mathbb{R}$, M has full rank:*

$$\operatorname{rank}(M) = 8,$$

and

$$\det(M) = -3 \neq 0.$$

In particular, M is invertible in $\operatorname{Mat}_{8\times 8}(\mathbb{R})$, and T is a linear automorphism of $\mathbb{R}^8$.

Proof. A direct computation (for example, by Gaussian elimination or with a computer algebra system) shows that the rows of M are linearly independent over $\mathbb{R}$. Thus $\operatorname{rank}(M) = 8$, which forces $\det(M) \neq 0$. An explicit determinant computation yields $\det(M) = -3$. □

Corollary 5. *Viewed as a matrix over $\mathbb{F}_2$, one has*

$$\det(M) \equiv -3 \equiv 1 \pmod 2,$$

so M is also invertible in $\operatorname{Mat}_{8\times 8}(\mathbb{F}_2)$. Hence M defines an automorphism of the 8-dimensional vector space $\mathbb{F}_2^8$:

$$M \in \mathrm{GL}(8, 2).$$

10.4 Spectral Properties Over $\mathbb{R}$ and $\mathbb{C}$

Over $\mathbb{R}$ (or $\mathbb{C}$), the characteristic polynomial of M is

$$\chi_M(x) = \det(xI_8 - M) = x^8 - 8x^5 - 38x^4 - 48x^3 - 24x^2 - 8x - 3.$$

This polynomial factors as

$$\chi_M(x) = (x-3)(x+1)^3\left(x^4 + 6x^2 + 1\right).$$

Proposition 6. *The eigenvalues of M over $\mathbb{C}$ are:*

$$\lambda_1 = 3, \quad \lambda_2 = -1 \text{ (with algebraic multiplicity 3)},$$

and the four roots of

$$x^4 + 6x^2 + 1 = 0,$$

which are purely complex conjugate pairs. The spectral radius of M is

$$\rho(M) = 3.$$

Proof (sketch). The factorization of $\chi_M(x)$ follows from a direct computation of $\det(xI_8 - M)$. The eigenvalues are precisely the roots of $\chi_M(x)$, with algebraic multiplicities given by the exponents of the linear factors in the factorization. The spectral radius is the maximum of the absolute values of the eigenvalues, which is $|3| = 3$. □

The vector **1** is an eigenvector for the eigenvalue 3, since

$$M\mathbf{1} = 3\mathbf{1}.$$

A computation of the dimensions of the eigenspaces shows that the geometric multiplicities match the algebraic multiplicities for all eigenvalues, so we obtain:

Proposition 7. *Over $\mathbb{C}$, the matrix M is diagonalizable. That is, there exists an invertible matrix $P \in \mathrm{GL}(8, \mathbb{C})$ such that*

$$P^{-1}MP = \mathrm{diag}(\lambda_1, \dots, \lambda_8),$$

where $\lambda_1, \dots, \lambda_8$ are the eigenvalues of M (listed with multiplicity).

Equivalently, the minimal polynomial of M over $\mathbb{C}$ is

$$\mu_M(x) = (x-3)(x+1)\left(x^4 + 6x^2 + 1\right),$$

each irreducible factor appearing with exponent 1.

10.5 Eigenstructure Over $\mathbb{F}_2$

Over the field $\mathbb{F}_2$, one can view M as a linear automorphism of $\mathbb{F}_2^8$. For example, the all-ones vector

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{F}_2^8$$

satisfies

$$M\mathbf{1} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} \equiv \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \pmod 2,$$

so $\mathbf{1}$ is an eigenvector with eigenvalue $1 \in \mathbb{F}_2$. Thus $\mathbf{1}$ spans a one-dimensional invariant subspace of $\mathbb{F}_2^8$ under M.

More generally, one may study the Jordan normal form of M over an algebraic closure of $\mathbb{F}_2$, or equivalently the factorization of $\chi_M(x)$ modulo 2. This describes the cycle structure of the permutation induced by M on the nonzero vectors of $\mathbb{F}_2^8$.

10.6 Graph-Theoretic Interpretation

Because all entries of M are 0 or 1, the matrix M can be interpreted as the adjacency matrix of a directed graph on the vertex set

$$\{1, 2, \dots, 8\}.$$

Definition 17. Define a directed graph G with vertex set $V(G) = \{1, \dots, 8\}$ and directed edges

$$i \to j \iff M_{ij} = 1.$$

Since every row and every column of M contains exactly three entries equal to 1, every vertex of G has out-degree 3 and in-degree 3. Thus G is a 3-regular directed graph (or, if we ignore edge orientation, a 3-regular simple graph, provided no symmetric pairs of edges yield multiple undirected edges).

From this viewpoint, the linear operator $T(x) = Mx$ on $\mathbb{R}^8$ (or $\mathbb{C}^8$) encodes walks on the graph G: the (i, j)-entry of M^k counts, for each $k \in \mathbb{N}$, the number of directed walks of length k from vertex i to vertex j.

10.7 Summary of Linear-Algebraic Features

To summarize the main linear-algebraic properties of the matrix

$$M = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

we note:

- M is a $(0,1)$-matrix with constant row sum and column sum equal to 3.
- $\operatorname{rank}(M) = 8$ and $\det(M) = -3$, so M is invertible over $\mathbb{R}$ and over $\mathbb{F}_2$, i.e. $M \in GL(8, \mathbb{R})$ and $M \in GL(8, 2)$.
- Over $\mathbb{C}$, the characteristic polynomial is $\chi_M(x) = (x-3)(x+1)^3(x^4+6x^2+1)$, and M is diagonalizable.
- Over $\mathbb{F}_2$, the matrix defines a linear automorphism of the vector space $\mathbb{F}_2^8$; in particular the all-ones vector is a fixed nonzero vector.
- Interpreted as an adjacency matrix, M encodes a 3-regular directed graph on 8 vertices.