

# Fundamentos de Organización de Datos

*Hashing*

# Dispersión de Archivos

- ✓ Técnica para generar una dirección base única para una clave dada.
- ✓ Convierte la clave en un número aleatorio, que luego sirve para determinar dónde se almacena la clave.
- ✓ Utiliza una función de dispersión para mapear cada clave con una dirección física de almacenamiento.
- ✓ Utilizada cuando se requiere acceso rápido por clave.

# Tipos de Dispersión

## **Direccionamiento estático**

El espacio disponible para dispersar los registros del archivo está fijado previamente.

## **Direccionamiento dinámico**

El espacio disponible para dispersar los registros del archivo aumenta o disminuye en función de las necesidades.

# Parámetros a considerar

Parámetros que influyen sobre el desempeño del ambiente de dispersión:

- ✓ Capacidad de almacenamiento de cada dirección
- ✓ Densidad de empaquetamiento
- ✓ Función de hash
- ✓ Método de tratamiento de desbordes

# Dispersión de Archivos

## **Función de dispersión**

Caja negra que a partir de una clave genera la dirección física donde debe almacenarse el registro.

## **Colisión**

Situación en la que un registro es asignado, por función de dispersión, a una dirección que ya posee uno o más registros.

# Dispersión de Archivos

## Desborde

Situación en la cual una clave carece de lugar en la dirección asignada por la función de dispersión.

## Densidad de empaquetamiento

Relación entre el espacio disponible para el archivo de datos y la cantidad de registros que integran el mismo.

$$DE = \text{número de registros} / \text{espacio Total}$$

# Dispersión de Archivos

Aunque la función de dispersión sea eficiente y la densidad de empaquetamiento sea baja, es probable que ocurran **desbordes**.

Métodos aplicables para resolver colisiones con desborde en *dispersión estática*:

- ✓ **Saturación progresiva**
- ✓ **Saturación progresiva encadenada**
- ✓ **Saturación progresiva con área de desborde por separado**
- ✓ **Dispersión doble**

**No vemos ejercicios prácticos de hashing estático**



*Hashing Extensible*



# Técnica de resoluciones: Hashing Extensible

Ejemplo:

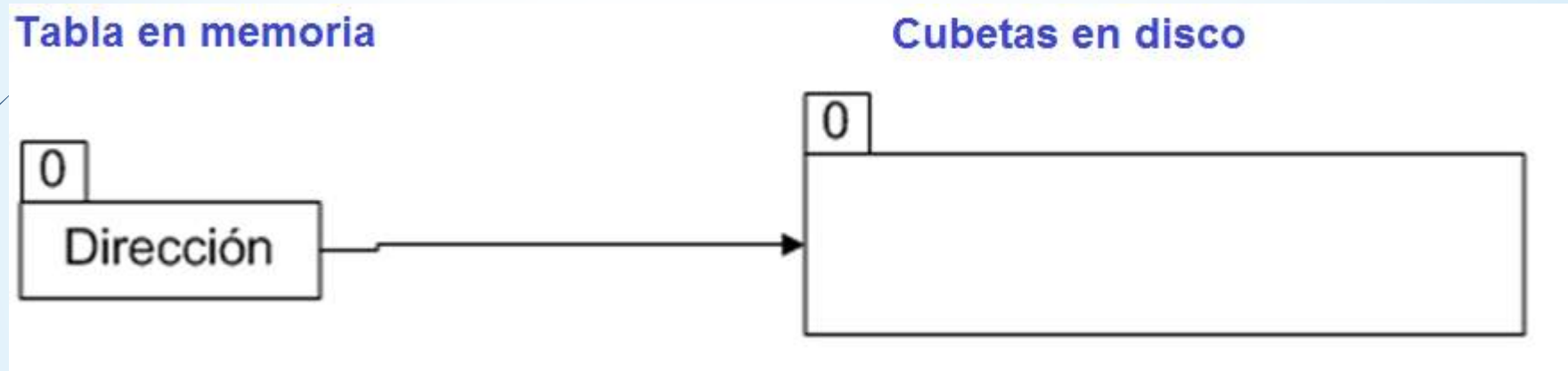
- Función de dispersión: Retorna 32 bits.
- Capacidad para 2 registros por dirección.
- Se van a dispersar 10 claves en total.

# Hashing Extensible

Clave	$f(\text{clave})$
Alfa	00.....1001
Beta	00.....0100
Gamma	00.....0010
Delta	00.....1111
Epsilon	00.....0000
Rho	00.....1011
Pi	00.....0110
Tau	00.....1101
Psi	00.....0001
Omega	00.....0111

# Hashing Extensible

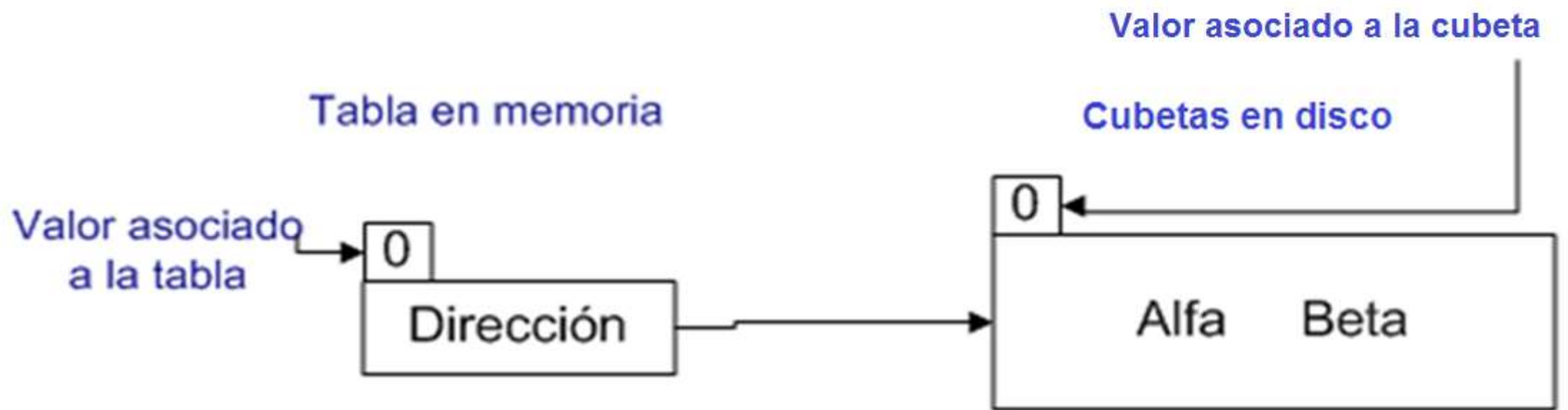
Estado inicial del archivo:



El número cero sobre la tabla indica que no es necesario ningún bit de la secuencia obtenida por la función de dispersión.

# Inserción de claves

Clave	$f(\text{clave})$
Alfa	00.....1001
Beta	00.....0100



## Inserción de claves - Desborde

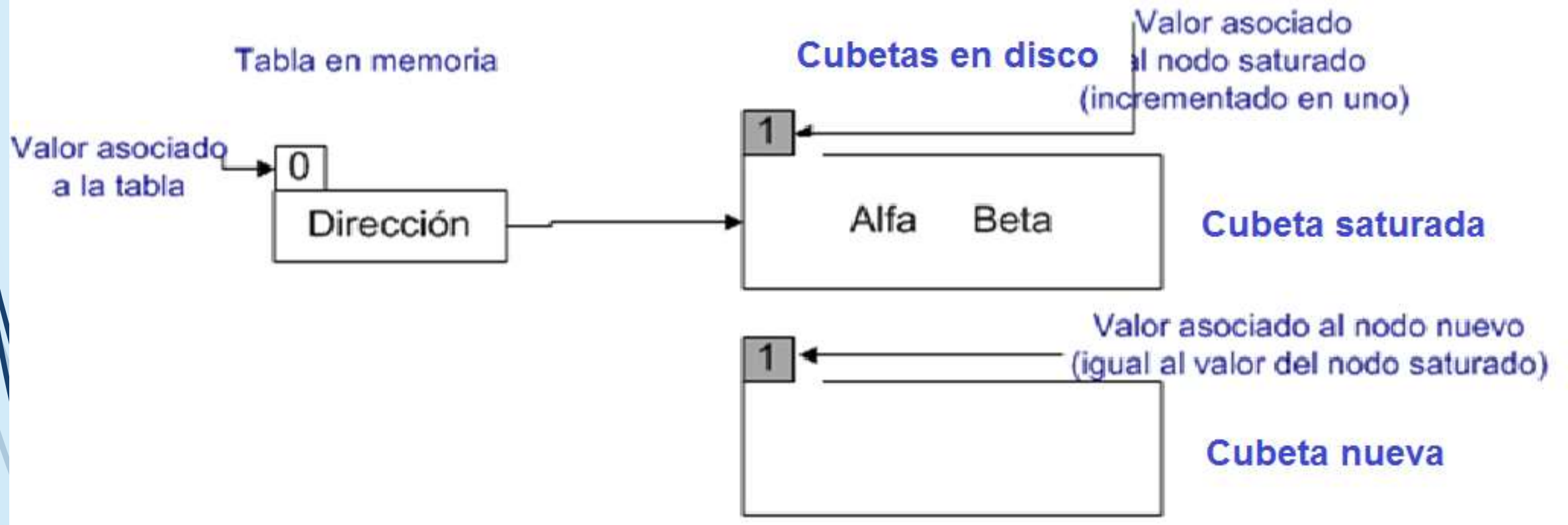
Clave	$f(\text{clave})$
Gamma	00.....0010

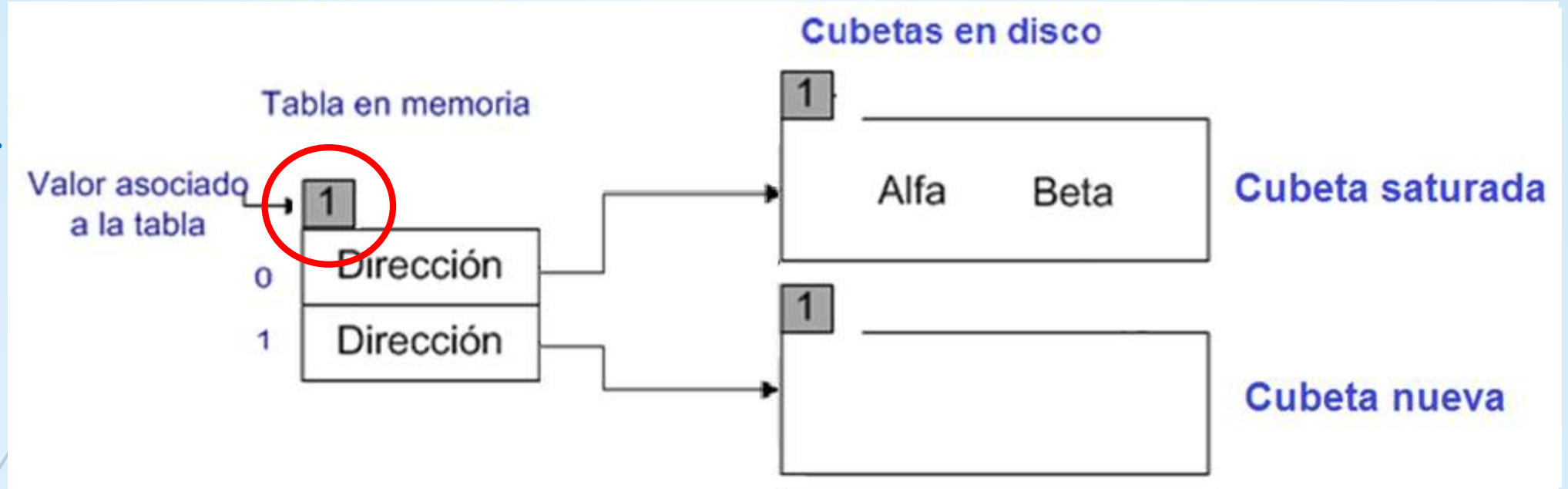
La inserción de Gamma produce **desborde**.

- 1) Se incrementa en uno el valor asociado a la cubeta saturada.
- 2) Se genera una nueva cubeta con el mismo valor asociado a la cubeta saturada.

# Inserción de Gamma

Clave	$f(\text{clave})$
Gamma	00.....0010





Se compara el valor de la cubeta con el valor asociado a la tabla -> El primero es mayor que el segundo.

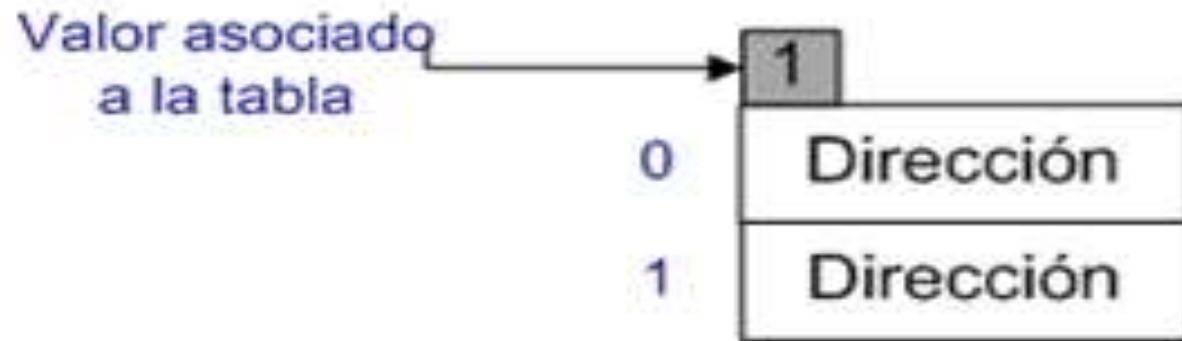
La tabla **no** dispone de entradas suficientes para direccionar a la nueva cubeta.

La tabla tiene una celda única, y como se dispone ahora de dos nodos, **hace falta generar más direcciones**.

La cantidad de celdas de la tabla se **duplica** y el valor asociado a la tabla **se incrementa en uno**.



### Tabla en memoria

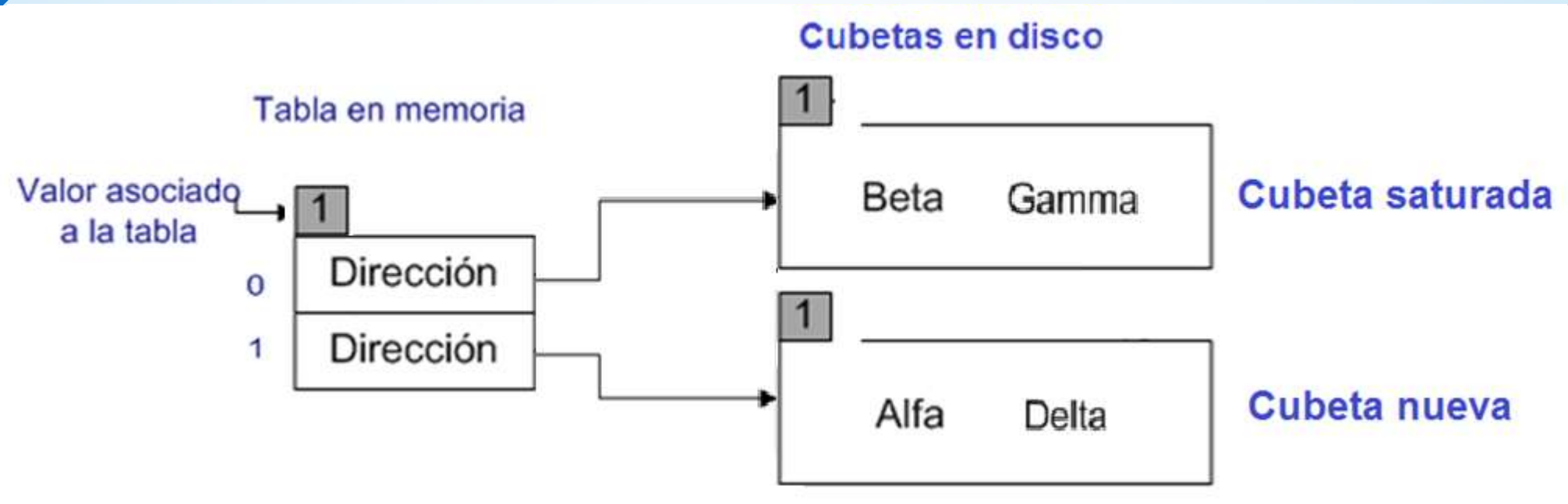


El valor asociado a la tabla indica la cantidad de bits que es necesario tomar de la función de hash.

La primera celda de la tabla direcciona a la cubeta saturada, y la nueva celda apunta a la nueva cubeta generada.



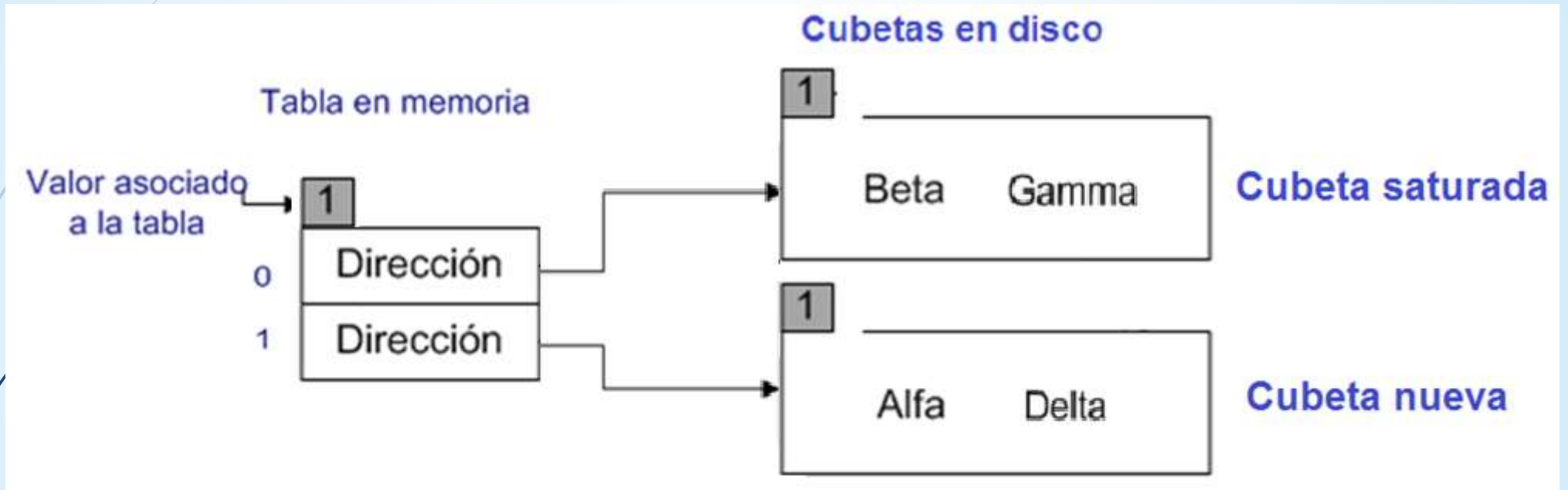
# Se redispersan las claves involucradas:



Clave	$f(\text{clave})$
Alfa	00.....1001
Beta	00.....0100
Gamma	00.....0010
<b>Delta</b>	<b>00.....1111</b>

# Inserción de Epsilon

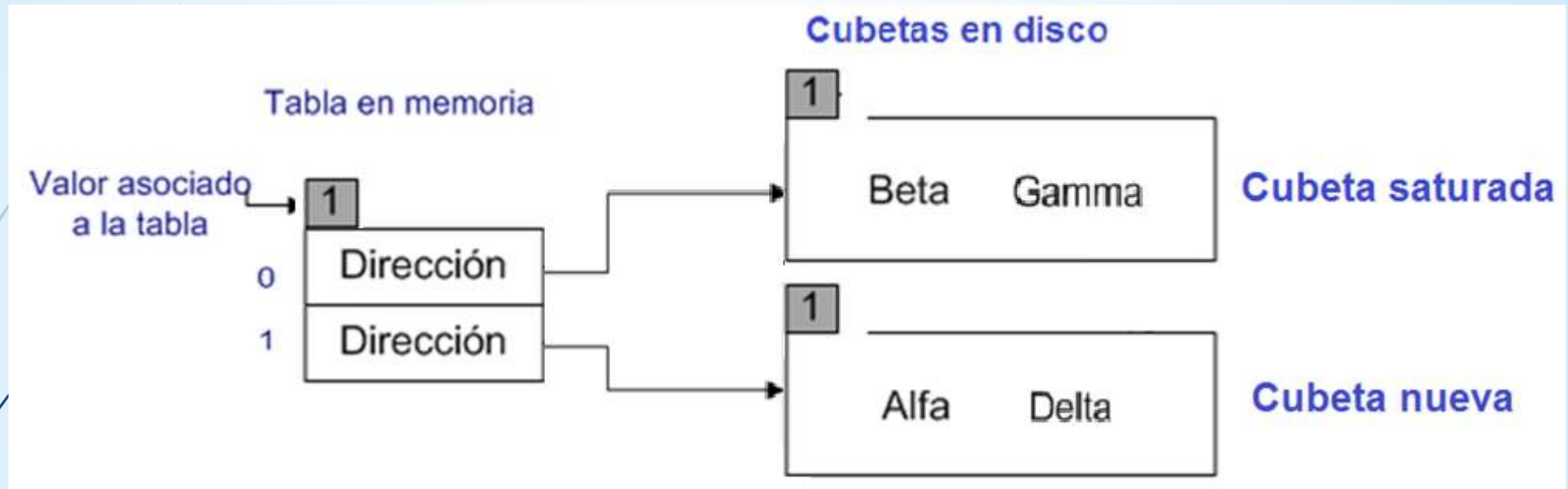
Clave	f(clave)
Epsilon	00.....0000



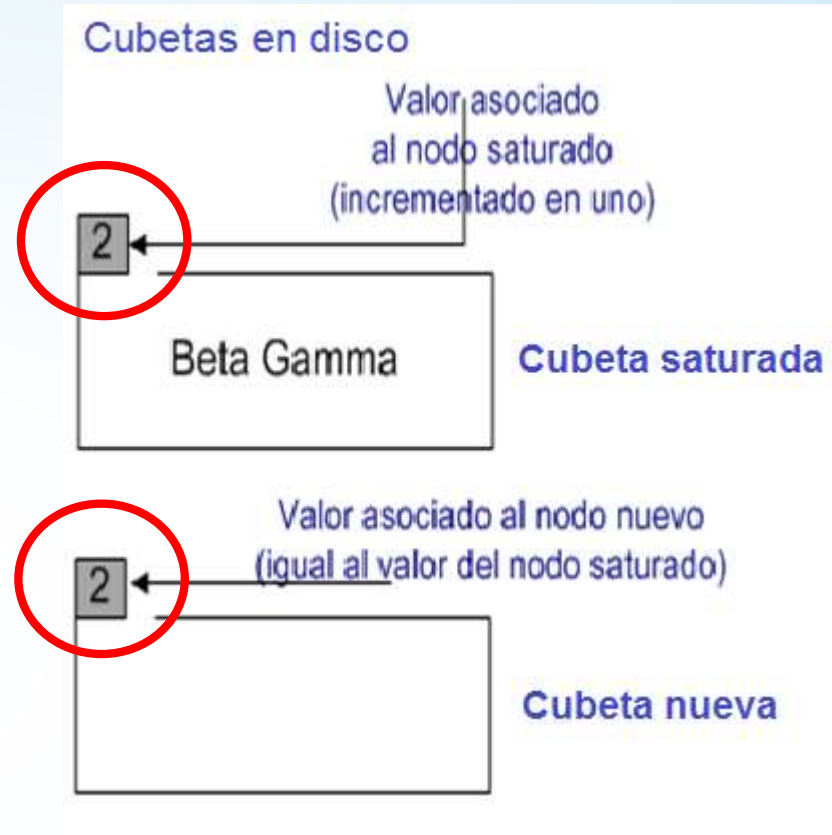
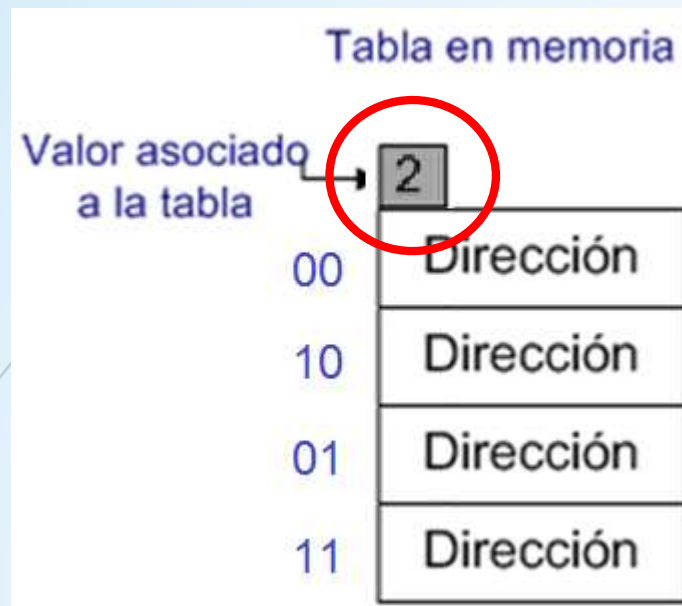
Epsilon debe ser almacenado en la cubeta asociada a la celda 0 de la tabla. La misma se encuentra completa lo que genera un nuevo **desborde**.

# Inserción de Epsilon

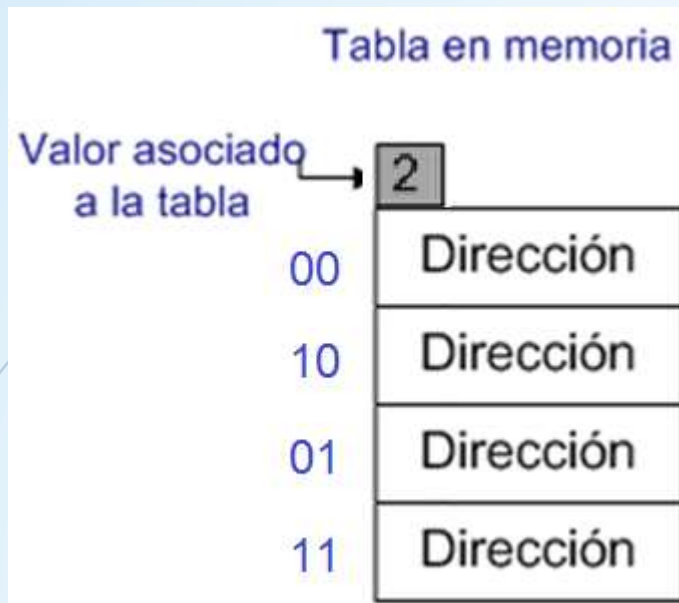
Clave	f(clave)
Epsilon	00.....0000



Al no disponer de celdas suficientes en la tabla en memoria principal, se **duplica** el espacio disponible, que a partir de este momento necesita 2 bits de la función de hash para poder direccionar un registro.



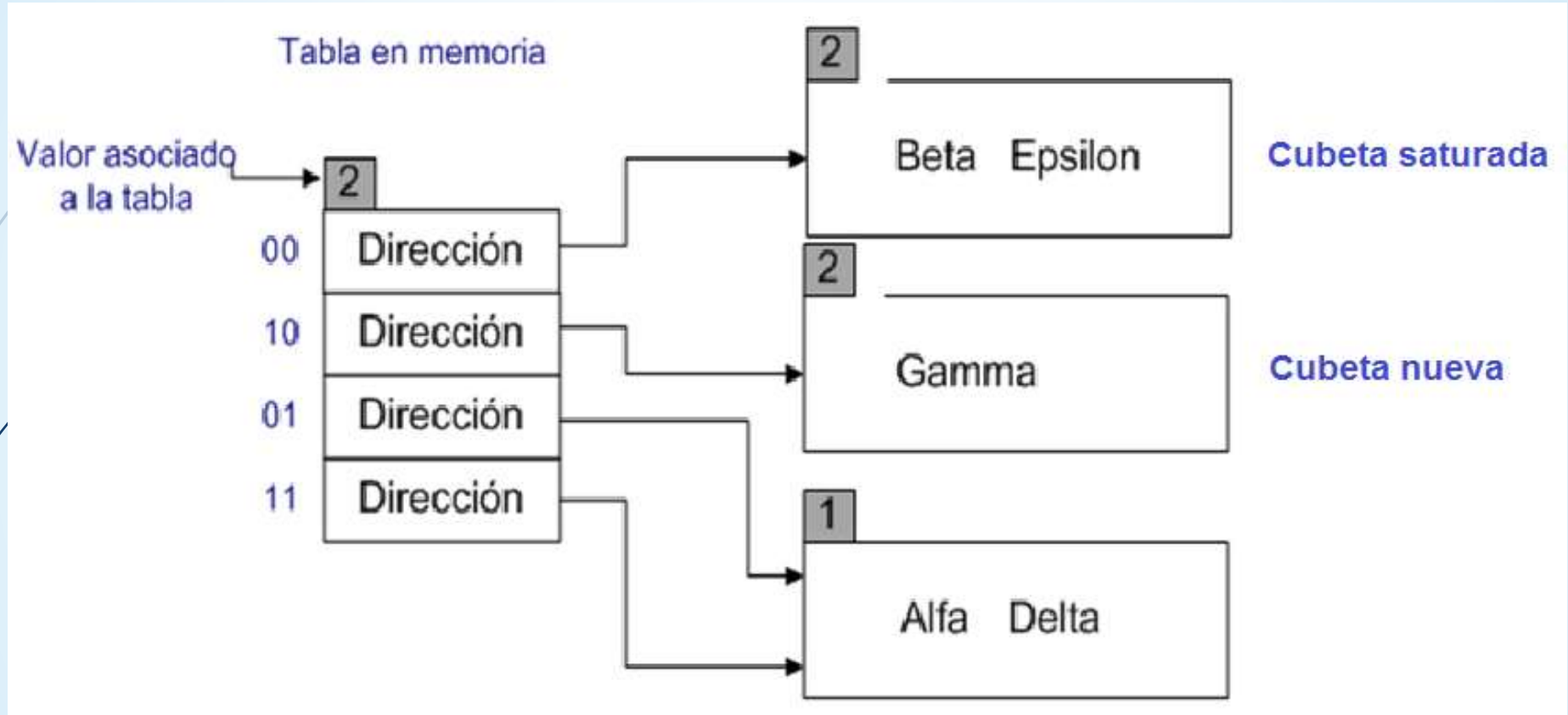
La celda de referencia 00 contiene la dirección de la cubeta saturada, mientras que la celda de referencia 10 contiene la dirección de la nueva cubeta.



Se redispersan **solamente** las claves de las cubetas involucradas:

Clave	f(clave)
Beta	00.....0100
Gamma	00.....0010
Epsilon	00.....0000

# Estado final luego de insertar Epsilon



Clave	f(clave)
Rho	00.....1011



## Insertión de Rho

Su dirección de almacenamiento corresponde a la cubeta asociada a la celda 11 → **Está completa**

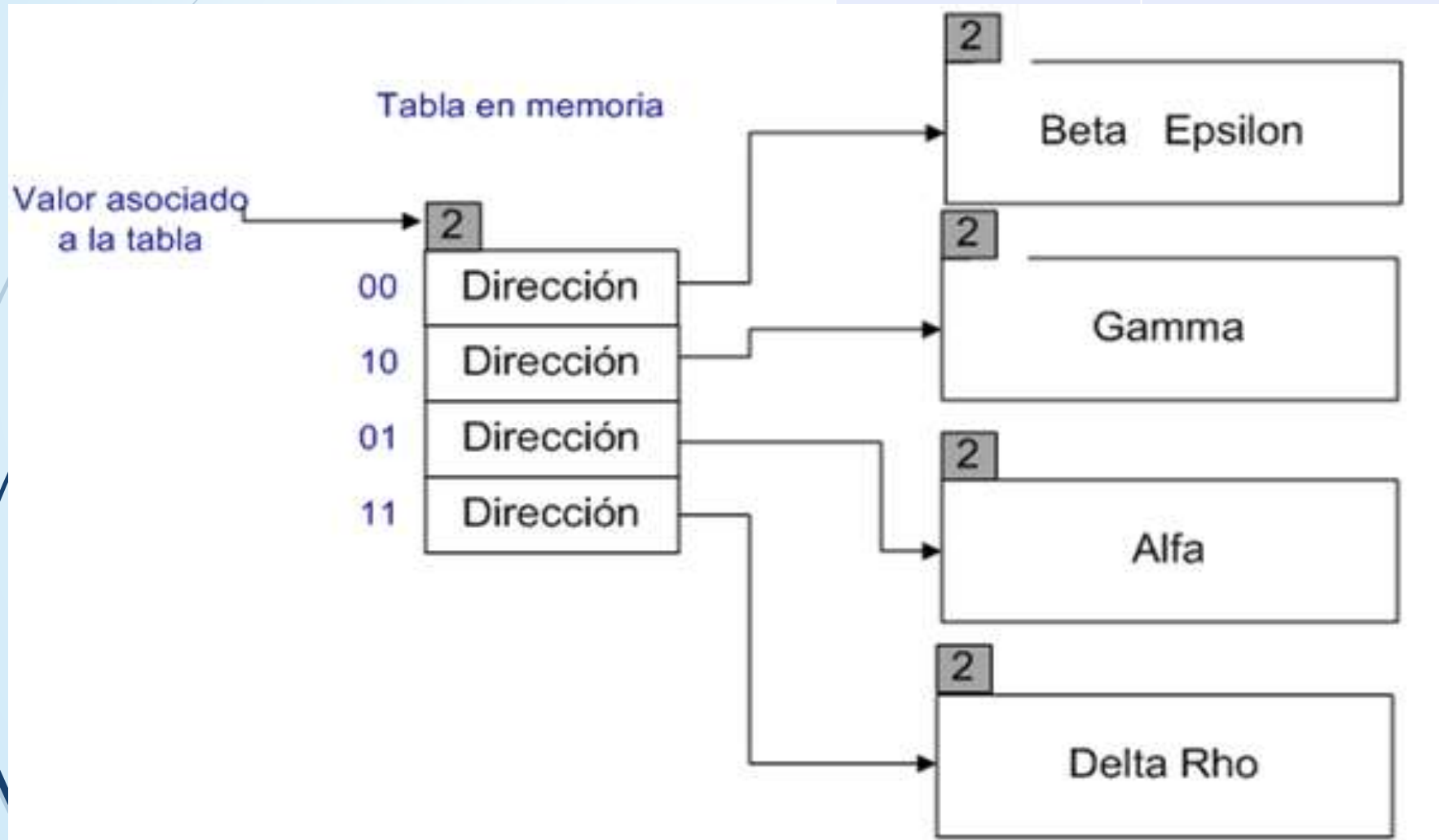
Se genera desborde y se crea una nueva cubeta.

El valor asociado a ambas cubetas coincide con el valor asociado a la tabla en memoria. Por lo tanto:

**La tabla posee direcciones suficientes para direccionar a la nueva cubeta y la cantidad de celdas NO debe ser duplicada!**

# Insertión de Rho

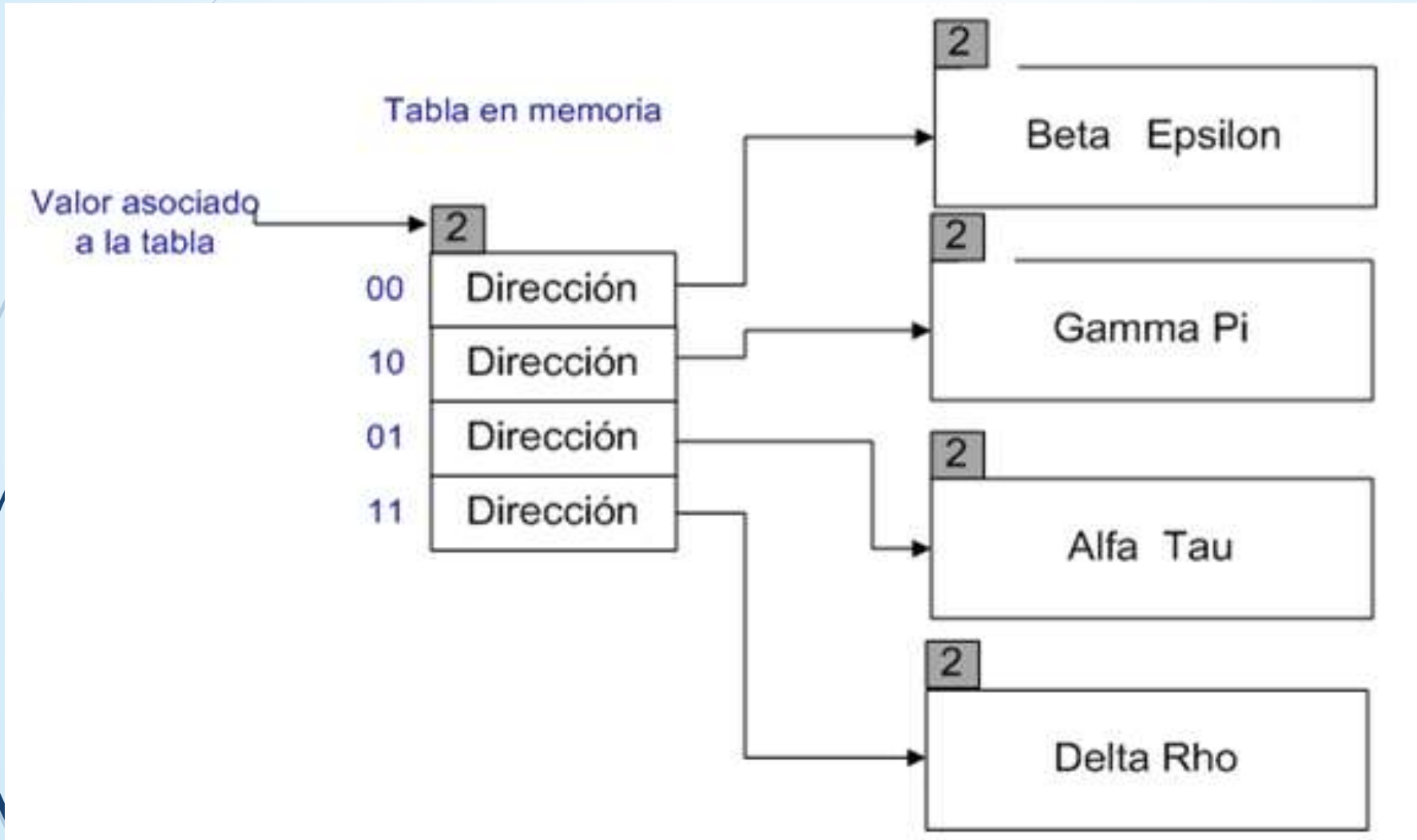
Clave	f(clave)
Alfa	00.....1001
Delta	00.....1111
Rho	00.....1011





# Inserción de Pi y Tau

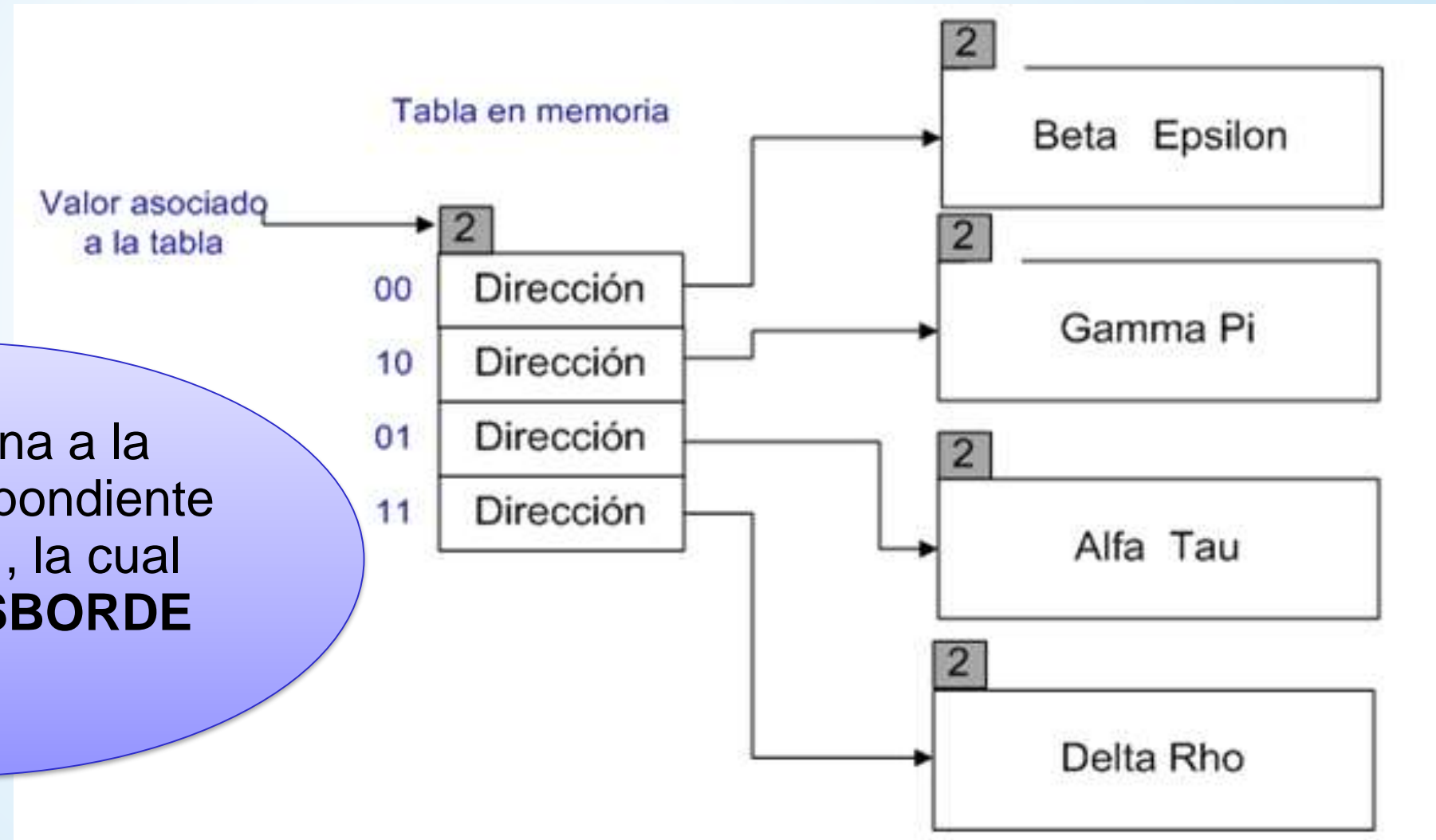
Clave	f(clave)
Pi	00.....0110
Tau	00.....1101



# Inserción de Psi

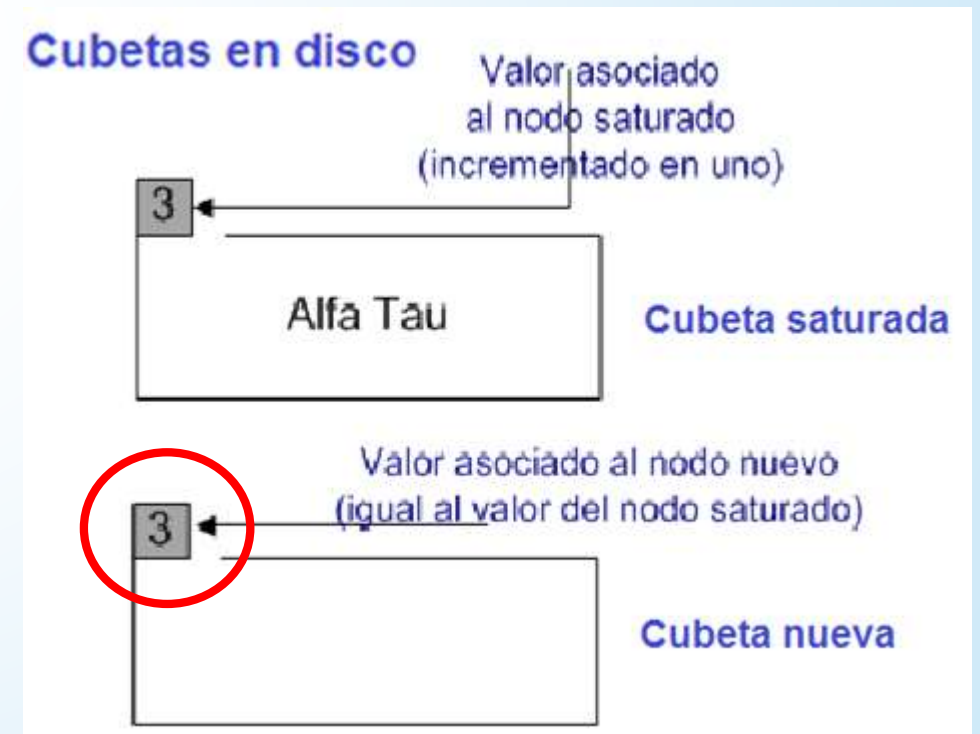
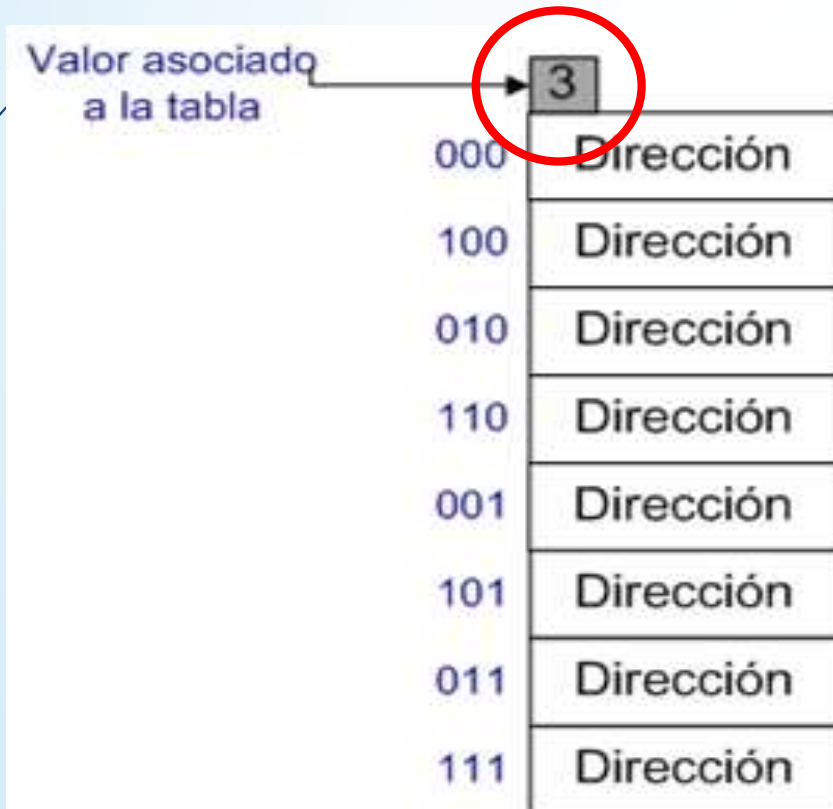
Clave	f(clave)
Psi	00.....0001

Se direcciona a la cubeta correspondiente a la celda 01, la cual produce **DESBORDE**

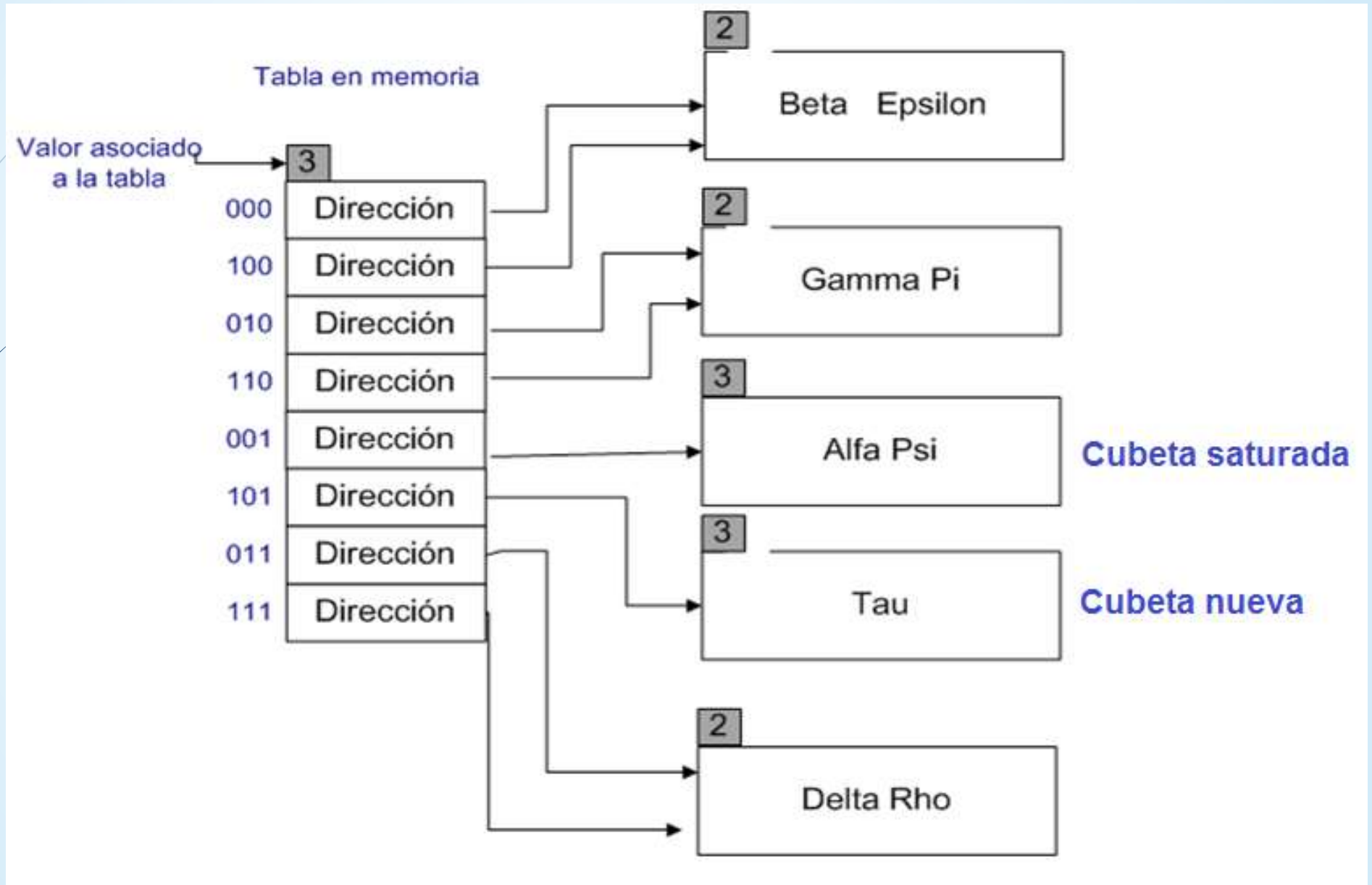


## Pasos:

1. Incrementar en uno el valor asociado al nodo con saturación.
2. Crear una nueva cubeta
3. Como el valor de la cubeta es **mayor** al valor de la tabla, se debe **duplicar** la tabla e incrementar su valor.



# Estado final luego de insertar Psi

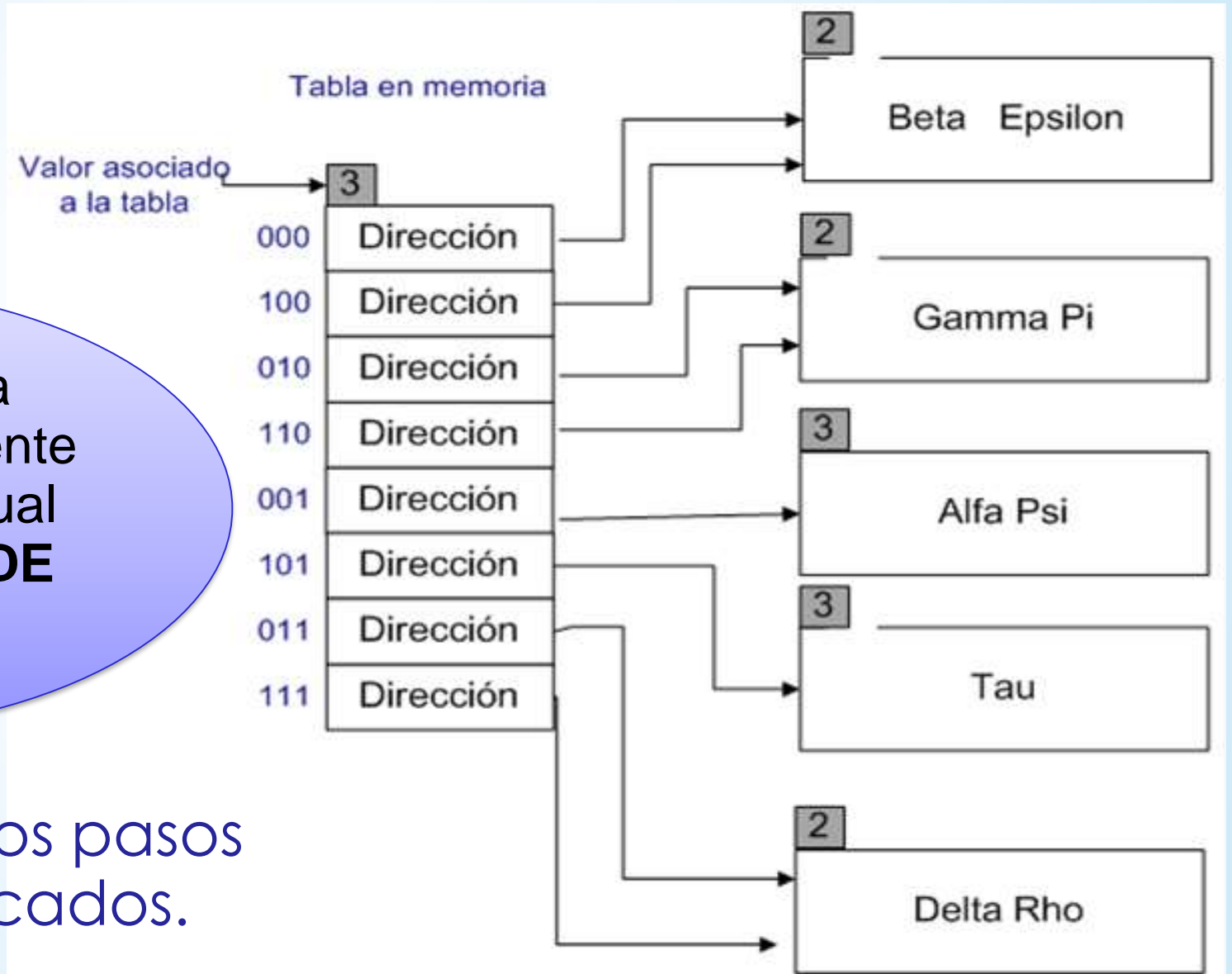


# Inserción de Omega

Clave	f(clave)
Omega	00.....0111

Se direcciona a la cubeta correspondiente a la celda 111, la cual produce **DESBORDE**

Se deben aplicar los pasos previamente explicados.



# Estado final

Clave	f (clave)
Rho	00.....1011
Delta	00.....1111
Omega	00.....0111

