

Capítulo 3: EVALUACIÓN DEL PARALELISMO Y LONGITUD VECTORIAL

En este capítulo se presenta la evaluación del paralelismo \parallel y longitud vectorial lv para las sentencias a partir del grafo de dependencias del bucle.

En la sección 1 se evalúa el paralelismo de bucles con grafos de dependencias acíclicos o cíclicos que no incluyen sentencias estructuradas condicionales. También se considera que todas las sentencias del bucle poseen el mismo tiempo de ejecución. En la sección 2 se evalúa la longitud vectorial con que pueden ejecutarse las sentencias de un bucle.

En la sección 3 se considera la influencia del tiempo de ejecución de las sentencias en el cálculo del paralelismo del bucle. En la sección 4 se considera la existencia de sentencias condicionales y su efecto en la evaluación del paralelismo.

El conocimiento del paralelismo es una buena medida que permite determinar la eficiencia de los procesos de paralelización descritos en este trabajo, permitiendo así su comparación.

3.1 PARALELISMO DE UN BUCLE

En esta sección se evalúa el paralelismo tanto para grafos acíclicos como cíclicos que representan las relaciones de dependencia que existen entre las n sentencias incluidas en un bucle que ejecuta N iteraciones. No se considera la existencia de sentencias estructuradas condicionales.

Se define *paralelismo* de un bucle \parallel como el número medio de procesadores activos que ejecutan diferentes instancias de las sentencias del bucle, dado un reparto de tareas óptimo y suponiendo costes de sincronización y reparto de tareas nulos.

Tal como se ha descrito en la sección 2.2, dada una relación de dependencia $S_i \delta S_j$ entre dos sentencias S_i y S_j de un bucle, se tiene que la ejecución de la iteración k de la sentencia S_i permite la ejecución de la iteración $k+d_{ij}$ de la sentencia S_j . Esta relación de precedencia implica que, el número de iteraciones iniciales de cada una de las sentencias S_j del bucle que pueden ser ejecutadas, sin violar ninguna de las relaciones de dependencia en las que S_j esté involucrada, viene dado por la menor de las distancias d_{ij} asociadas a los arcos $d_{ij} \in E$ que entran al nodo $S_j \in V$ del grafo. Esto es así dado que la primera iteración i_0 de la sentencia fuente de la dependencia $S_i \in V$ libera la ejecución de la iteración $(i_0 + d_{ij})$ de la sentencia destino S_j . Por lo tanto, todas las iteraciones anteriores a esta podrán ser inicialmente ejecutadas.

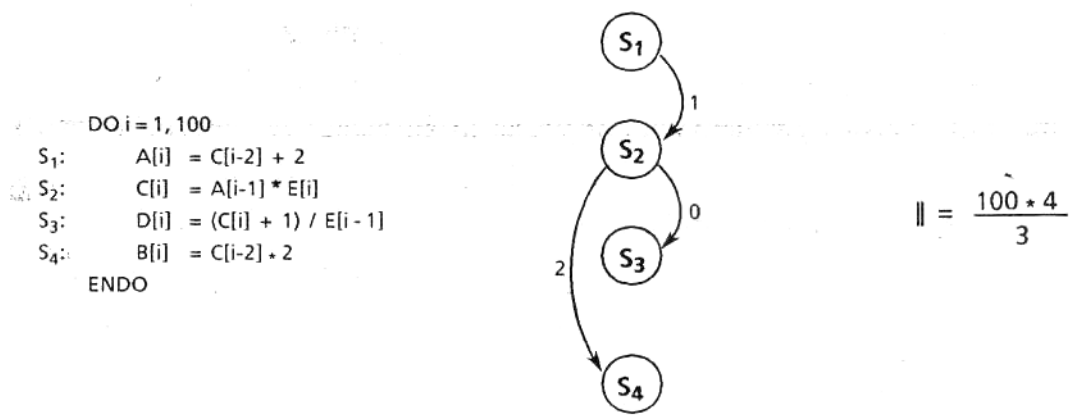
3.1.1 Bucles con GDA

La figura 3.1(a) se muestra un bucle secuencial caracterizado por un grafo de dependencias acíclico (GDA) en el que se observan dos cadenas de dependencia: $\{d_{12}, d_{24}\}$ y $\{d_{12}, d_{23}\}$, ambas con una longitud asociada de 3. Tal como muestra la ejecución desarrollada del bucle de la figura 3.1(b), en la cual se representan en vertical las sentencias del bucle y en horizontal las iteraciones, es posible ejecutar en un primer ciclo todas las iteraciones de S_1 , la primera iteración de S_2 y las dos primeras iteraciones de S_4 en paralelo, dado que no existe ninguna relación de dependencia que impida su ejecución. En un segundo ciclo es posible ejecutar el resto de iteraciones de S_2 , la primera iteración de S_3 y la tercera iteración de S_4 en paralelo, dado que se han liberado las relaciones de dependencia que impedían su ejecución. Por último en un tercer ciclo es posible ejecutar el resto de iteraciones de S_3 y S_4 en

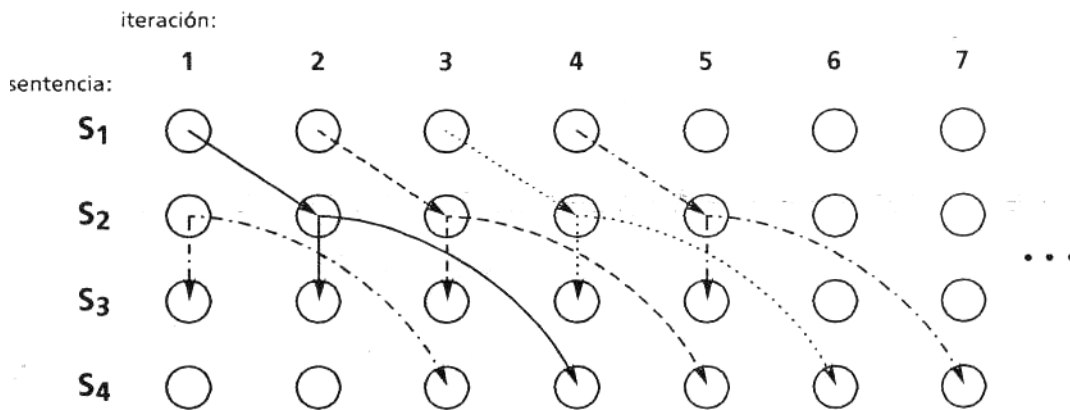
paralelo. Por lo tanto el paralelismo del bucle, evaluado como el número total de instanciaciones de sentencias a ejecutar dividido por el tiempo requerido para su ejecución, es

$$\parallel = \frac{100 \cdot 4}{3} = 133, \bar{3} .$$

Observar que el paralelismo no depende de las distancias que acompañan a los arcos del grafo de dependencias, sino únicamente del orden impuesto por las relaciones de dependencia entre sentencias.



(a)



(b)

Figura 3.1: Cálculo del paralelismo para un bucle con GDA.

En general, para un bucle caracterizado por un GDA, el paralelismo puede ser evaluado como el número total de instanciaciones de todas las sentencias del bucle dividido por el tiempo de ejecutar la cadena de dependencias más larga del grafo, es decir,

3.1.2 Bucles una única recurrencia hamiltoniana

$$\parallel = \frac{N * n}{l(C)}$$

siendo C la cadena con mayor longitud asociada

$$l(C) \geq l(C_{ij}) \quad \forall C_{ij} \mid S_i, S_j \in V.$$

En la figura 3.2(a) se muestra un bucle secuencial y su GDC hamiltoniano. En la figura 3.2(b) se muestra parte de la ejecución secuencial del bucle desarrollada. Si denotamos con S_{ij} la instanciación de la sentencia S_i en la iteración j , se observa que inicialmente pueden ser ejecutadas S_{11} , S_{12} , S_{21} , S_{31} y S_{32} , dado que su ejecución no depende de la ejecución previa de otras sentencias. A partir de ellas se generan 5 cadenas de ejecución independientes, que determinan el paralelismo del bucle.

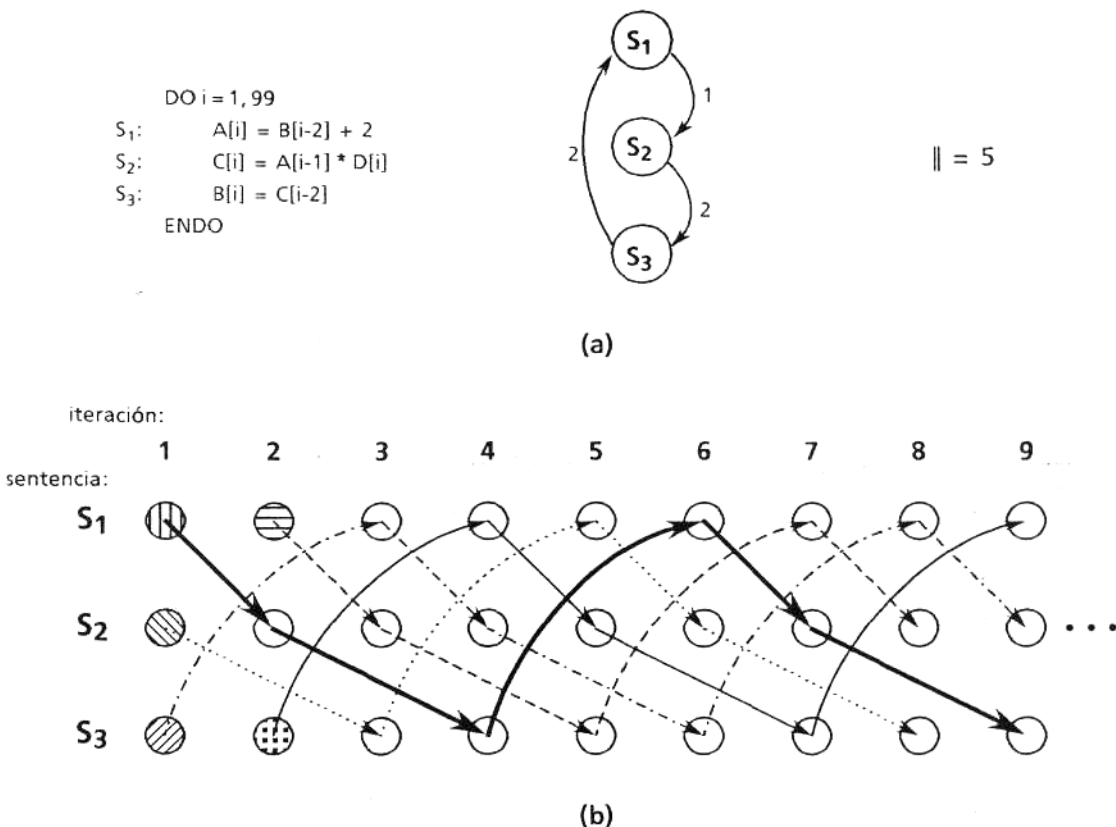


Figura 3.2: (a) Cálculo del paralelismo para un bucle con GDC y un único ciclo hamiltoniano y (b) ejecución desarrollada.

En general, dado un GDC con una única recurrencia R hamiltoniana, se tiene que el número de iteraciones de cada sentencia S_j que pueden ser inicialmente ejecutadas viene determinado por la distancia asociada al arco

que entra al nodo del grafo que la representa. Por lo tanto, inicialmente se pueden ejecutar $w(R)$ operaciones distintas. Dado que la ejecución de una sentencia libera a la sentencia que directamente depende de ella, se tiene que en cada instante de tiempo pueden ser ejecutadas $w(R)$ instancias de sentencias del bucle.

El paralelismo de un bucle que posee un GDC $G(V, E)$ caracterizado por poseer una única recurrencia R hamiltoniana viene dado por la suma de las distancias de los arcos que constituyen la recurrencia, es decir,

$$\| = w(R) = \sum_{d_{ij} \in R} d_{ij} .$$

El paralelismo obtenido con el razonamiento anterior también puede evaluarse basándose en la ejecución desarrollada del bucle y considerando que viene limitado por la ejecución de la cadena de dependencias más larga del grafo.

En el ejemplo de la figura 3.2 se puede observar que todas las cadenas de dependencia son iguales y quedan caracterizadas por un patrón cíclico que:

- en cada paso avanza $w(R)$ iteraciones, siendo R la recurrencia hamiltoniana del grafo;
- requiere para ser ejecutado un número de ciclos determinado por la cardinalidad de la recurrencia $|R|$;
- se repite un número de veces durante la ejecución de las N iteraciones dado por

$$\frac{N}{w(R)}$$

El tiempo necesario para ejecutar una de las cadenas de dependencia viene determinado por el tiempo necesario para ejecutar un paso por la recurrencia multiplicado por el número de pasos que se realizan durante la ejecución de las N iteraciones. Por otro lado, el número total de instancias de sentencias ejecutadas viene determinado por el producto del número total de iteraciones N a ejecutar y el número de sentencias n del bucle. Por lo tanto, el paralelismo del bucle queda determinado por

$$\| = \frac{n * N}{|R| * \frac{N}{w(R)}} = n * \frac{w(R)}{|R|} .$$

En el caso de existir una única recurrencia hamiltoniana se cumple que $|R| = n$ y por lo tanto,

$$\parallel = w(R) = \sum_{d_{ij} \in R} d_{ij} .$$

3.1.3 Bucles con GDC

El paralelismo de un bucle caracterizado por un GDC que posee varias recurrencias viene limitado por la recurrencia más restrictiva del grafo. A continuación se muestra cómo determinar cual es la recurrencia que determina el paralelismo del bucle.

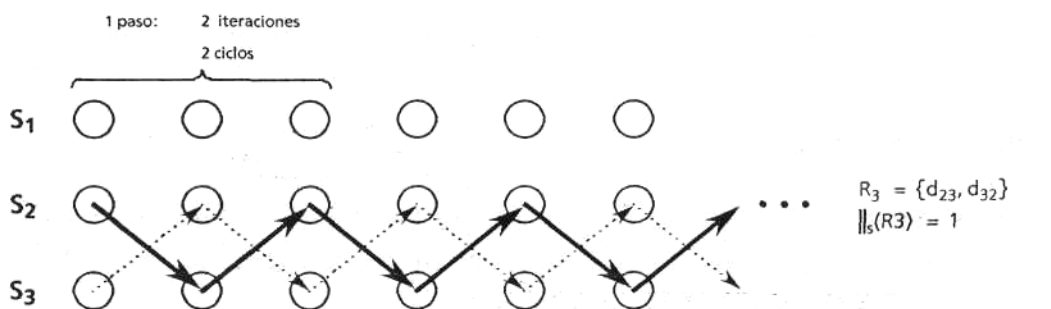
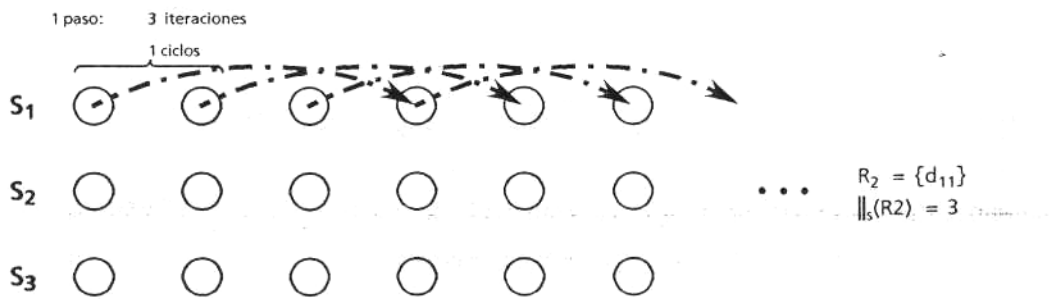
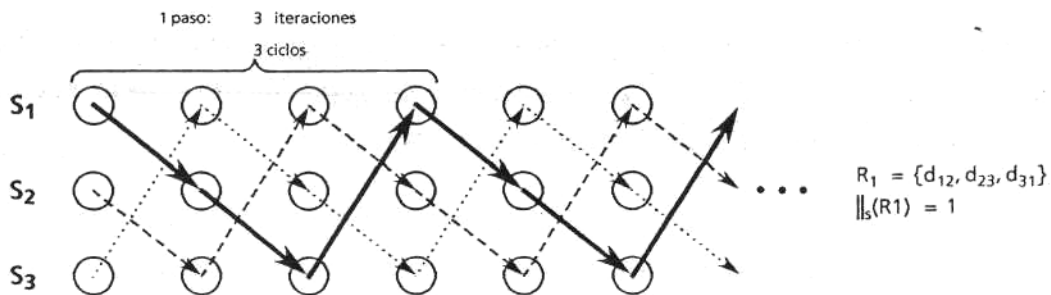
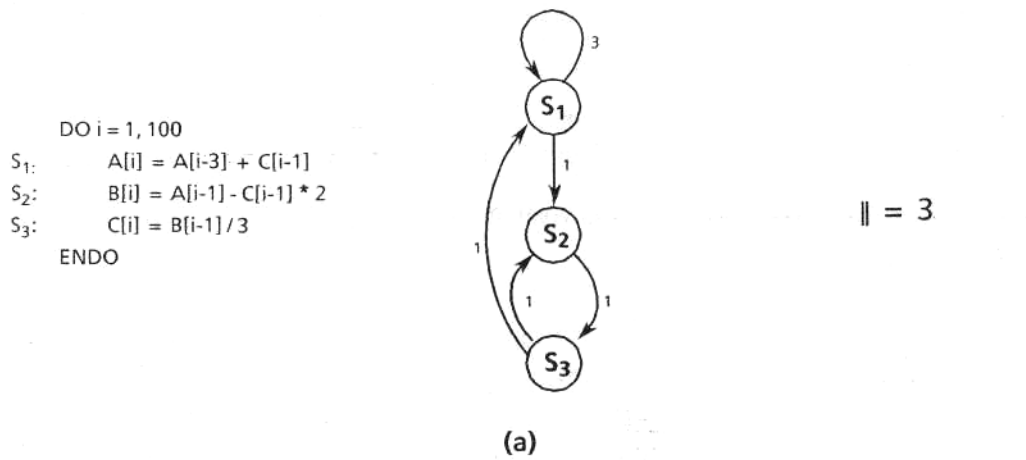
Por ejemplo, si se considera el bucle secuencial y grafo de dependencias asociado de la figura 3.3(a), se puede observar en su ejecución desarrollada (figura 3.3(b)) que cada cadena de dependencias avanza o cubre un determinado número de iteraciones del bucle en un determinado número de ciclos o unidades de tiempo. Así por ejemplo, la cadena de dependencias generada por la recurrencia $R_1 = \{d_{12}, d_{23}, d_{31}\}$ avanza 4 iteraciones en 3 ciclos. El número de iteraciones que cubre una cadena viene determinado por el peso de la recurrencia $w(R)$ que la genera. El tiempo requerido para realizar un avance viene determinado por la cardinalidad de la recurrencia $|R|$.

Se define *paralelismo por sentencia* \parallel_s como el número de iteraciones que avanza una cadena de dependencias dividido por el número de ciclos requeridos para realizar dicho avance, es decir,

$$\parallel_s(R) = \frac{w(R)}{|R|} .$$

Así por ejemplo, en el bucle de la figura 3.3, se observan tres recurrencias $R_1 = \{d_{12}, d_{23}, d_{31}\}$, $R_2 = \{d_{11}\}$ y $R_3 = \{d_{23}, d_{32}\}$ con paralelismos por sentencia asociados $\parallel_s(R_2) = 3$ y $\parallel_s(R_1) = \parallel_s(R_3) = 1$. El paralelismo del bucle \parallel queda limitado por aquellas recurrencias que poseen un paralelismo por sentencia menor.

Dado un GDC $G(V, E)$, se define B como el conjunto de recurrencias del grafo.



(b)

Figura 3.3: Cálculo del paralelismo para un bucle con GDC con varias recurrencias (a) y ejecución desarrollada (b).

Basándose en las expresiones del apartado anterior, para cada recurrencia $R \in B$, se tiene un paralelismo dado por

$$n \cdot \frac{w(R)}{|R|} = n \cdot \parallel_s(R).$$

Aquella recurrencia que posea un \parallel_s menor será la que limite el paralelismo \parallel del grafo, de manera que

$$\parallel = n \cdot \min_{R \in B} (\parallel_s(R)),$$

siendo n el número de sentencias del bucle.

Por lo tanto, el bucle del ejemplo anterior tiene un paralelismo asociado de 3, limitado por las recurrencias R_2 y R_3 .

Notar que el paralelismo evaluado según la expresión anterior puede dar un valor no entero. En este caso, la distribución de operaciones deberá seguir un patrón cíclico con paralelismo medio igual al evaluado.

Observar también que la expresión del paralelismo para GDA obtenida en el apartado 3.1.1 puede obtenerse de la anterior expresión considerando una recurrencia ficticia que incluyera el camino de dependencias más largo del grafo y cuyo peso asociado fuera el número de iteraciones N del bucle.

3.2 LONGITUD VECTORIAL DE SENTENCIAS EN BUCLES

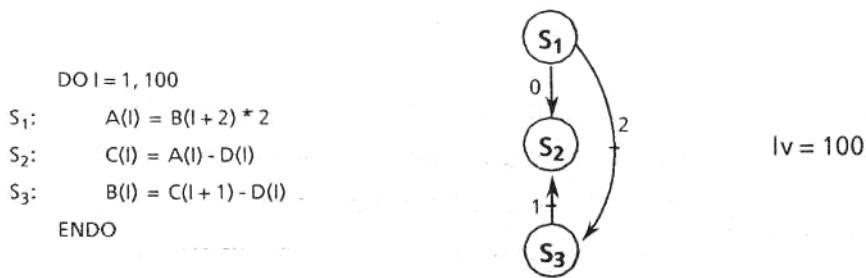
En esta sección se evalúa la longitud vectorial de cada una de las n sentencias que constituyen un bucle que ejecuta N iteraciones.

Se define *longitud vectorial* lv de una sentencia como el máximo número de iteraciones de dicha sentencia que pueden ser ejecutadas de forma vectorial.

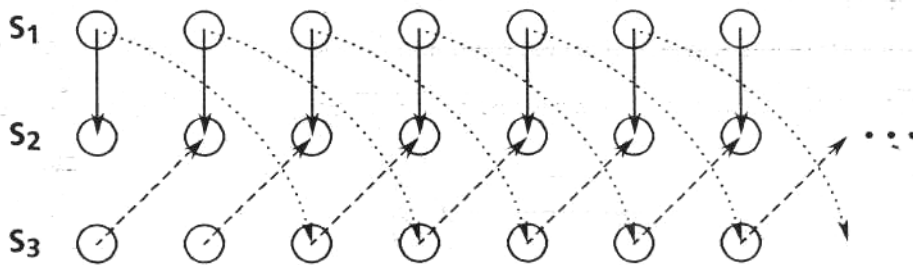
3.2.1 Bucles con GDA

Para el caso de grafos acíclicos $G(V, E)$, la longitud vectorial lv de las sentencias viene determinada por el número de iteraciones del bucle N , tal como se ha descrito en el método de distribución de bucles.

El orden en que las sentencias deben de ser ejecutadas viene determinado por las relaciones de dependencias entre ellas, pudiendo aplicar en este caso la técnica de reordenación de sentencias.



(a)



(b)

Figura 3.4: Cálculo de la lv de las sentencias de un bucle con GDA.

La figura 3.4(a) muestra un bucle y GDA asociado. La ejecución desarrollada de la figura 3.4(b) muestra que es posible ejecutar la sentencia S_i con $lv = 100$. Después de ella, al liberarse todas las relaciones de dependencia d_{13} podrá ejecutarse S_3 con la misma longitud vectorial lv . Su ejecución libera a su vez las dependencias d_{32} permitiendo por último la ejecución de S_2 con la misma longitud vectorial lv que las anteriores.

3.2.2 Bucles con GDC

A continuación se considera la evaluación de la longitud vectorial para sentencias incluidas en ciclos de dependencias, tanto para grafos hamiltonianos con una única recurrencia, en cuyo caso todas las sentencias llevan asociada la misma longitud vectorial, como para grafos en general, en cuyo caso cada sentencia S_i puede llevar asociada una longitud vectorial $lv(S_i)$ distinta.

Gratos con una única recurrencia hamiltoniana

Si se considera el grafo de dependencias y ejecución desarrollada para el bucle de la figura 3.5, se observa que únicamente existe una recurrencia que cubre todos los nodos del grafo. Si se suponen ejecutadas las instanciaciones sombreadas, se observa que el número de iteraciones de S_1 que pueden ser ejecutadas en forma vectorial viene limitado por la autodependencia $S_1 \delta S_1$ que se genera de forma indirecta a través de la recurrencia hamiltoniana, y que en este caso determina una longitud vectorial $lv = 5$. A partir de aquí es posible ejecutar de forma alternada las sentencias del bucle con dicha longitud vectorial en tantas pasadas como sean necesarias para cubrir las N iteraciones del bucle.

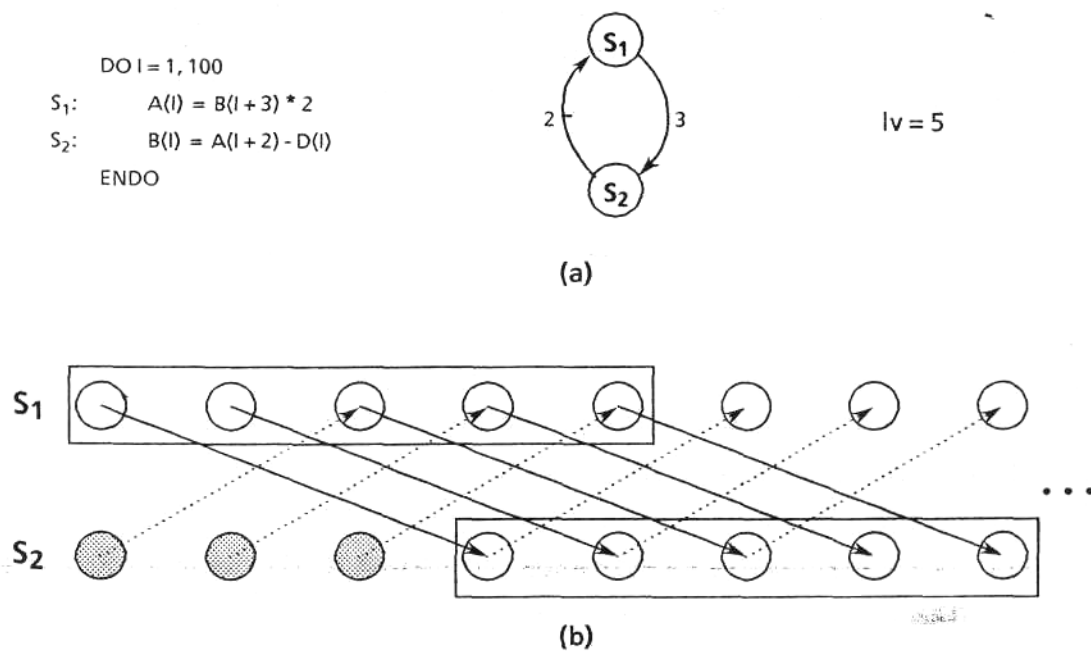


Figura 3.5: Cálculo de la lv para las sentencias de un bucle con GDC y una única recurrencia hamiltoniana.

La longitud vectorial de las sentencias que constituyen un bucle caracterizado por un GDC, con una única recurrencia hamiltoniana es tal que

$$lv \in \mathbb{N} \mid S_{i \ k} \delta S_{i \ k+lv}$$

siendo en este caso

$$lv = w(\mathbf{R}) = \sum_{d_{ij} \in \mathbf{R}} d_{ij}$$

Bucles con GDC en general

Si se considera un bucle caracterizado por un grafo de dependencias cíclico $G (V, E)$ con varias recurrencias, se tiene que la longitud vectorial de cada sentencia S_i viene determinada por el peso de la recurrencia más cerrada o restrictiva que la incluye, es decir,

$$lv(S_i) \in \mathbb{N} \mid S_{i,k} \delta S_{i,k+lv(S_i)}$$

y por lo tanto

$$lv(S_i) = \min_{R \mid S_i \in R} w(R).$$

Así por ejemplo, para el bucle de la figura 3.3(a) se tiene que S_1 se puede ejecutar con $lv(S_1) = 3$, limitada por las recurrencias R_1 y R_3 y S_2 y S_3 con $lv(S_2) = lv(S_3) = 2$, limitadas por la recurrencia R_2 .

3.3 INFLUENCIA DEL TIEMPO DE EJECUCIÓN

En esta sección se considera la influencia que ejerce el tiempo de ejecución de las sentencias en el cálculo del paralelismo del bucle.

El hecho de que una sentencia tarde m unidades de tiempo en ejecutarse equivale a replicar su nodo del grafo en m nodos con tiempo de ejecución unitario y encadenarlos con arcos de distancia 0. Las figuras 3.6(a) y 3.6(c) muestran dos ejemplos del mismo grafo de dependencias con duración distinta, de las sentencias. Cada nodo del grafo está etiquetado con el número de unidades de tiempo que tarda en ejecutarse. Observar que si el nodo que tarda más tiempo en ejecutarse no pertenece a la recurrencia que limita el paralelismo, este aumenta. Por el contrario, si pertenece a ella, entonces el paralelismo disminuye.

Con la finalidad de generalizar las expresiones de cálculo del paralelismo, se amplía el grafo de dependencias tal como se ha descrito en el apartado 2.2.5 de manera que cada nodo $S_k \in V$ del grafo $G (V, E)$ va etiquetado con el número de unidades de tiempo que la sentencia S_k requiere para ejecutarse.

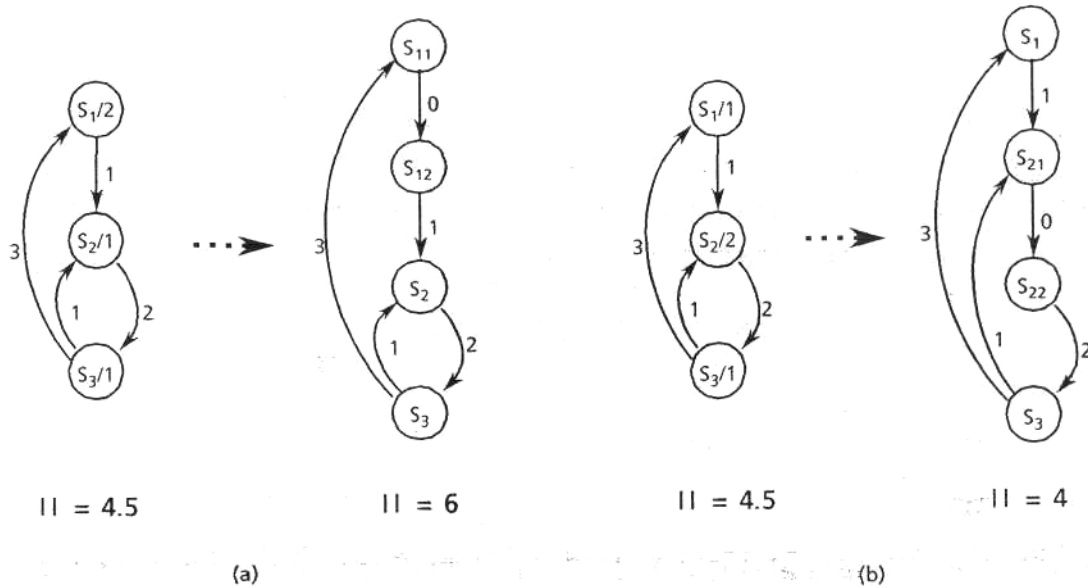


Figura 3.6: Influencia del tiempo de ejecución en el cálculo del paralelismo.

En este caso, el tiempo de ejecución de un camino del grafo C_{ij} viene determinado por su longitud

$$l(C_{ij}) = \sum_{S_k \in C_{ij}^*} S_k.$$

Se define *paralelismo por unidad de tiempo* para una determinada recurrencia $R \in B$ del grafo como

$$\parallel_t(R) = \frac{w(R)}{l(R)}$$

El paralelismo del bucle vendrá determinado por lo tanto por la recurrencia más cerrada del grafo, o sea,

$$\parallel = T \cdot \min_{R \in B} (\parallel_t(R))$$

siendo T el tiempo de ejecución total de una iteración del bucle

$$T = \sum_{S_k \in V} S_k.$$

3.4 SENTENCIAS CONDICIONALES

En esta sección se evalúa el paralelismo de bucles que incluyen sentencias estructuradas condicionales, de manera que pueden existir varias ramas posibles de ejecución en cada iteración del bucle, entendidas como el conjunto

de sentencias del bucle ejecutadas según los resultados de las condiciones evaluadas.

En el primer apartado se evalúan cotas para el paralelismo del bucle a partir de los paralelismos de cada una de las posibles ramas de ejecución y del paralelismo de recurrencias que puedan existir entre ellas. En el segundo apartado se estudia la variación del paralelismo en función de la probabilidad de cada una de las ramas de ejecución.

3.4.1 Cotas del paralelismo

En este apartado se obtienen cotas para el paralelismo máximo y mínimo del bucle considerando en primer lugar la evaluación del paralelismo para cada una de las posibles ramas de ejecución. Para ello se presenta el método que permite obtener el subgrafo de dependencias asociado a cada una de las ramas de ejecución.

En segundo lugar se considera la evaluación del paralelismo de las recurrencias que aparecen en el grafo de dependencias, considerando la secuencia de ejecución de las ramas condicionales que determina el paralelismo mínimo de dichas recurrencias.

3.4.1.1 Cota superior del paralelismo

La cota superior del paralelismo viene determinada por el máximo de los paralelismos de las posibles ramas de ejecución, evaluados a partir del subgrafo de dependencias asociado tal como se ha descrito en la sección 3.1.

Las relaciones de dependencia que aparecen en el subgrafo de dependencias asociado a una rama de ejecución son:

- (a) las relaciones de dependencia entre sentencias de dicha rama que originalmente ya existen en el grafo de dependencias del bucle;
- (b) las relaciones de dependencia debidas a efectos laterales de otros arcos del grafo original que afectan a sentencias no pertenecientes a la rama, pero que de forma indirecta, establecen una relación de dependencia entre sentencias de la rama considerada.

Efecto lateral de arcos incluidos en otras ramas

La figura 3.7(a) muestra un bucle que incluye una sentencia condicional del tipo IF-THEN. La figura 3.7(b) muestra el grafo de dependencias asociado. Si se supone que la condición evaluada en S_2 nunca va a cumplirse, entonces S_3 nunca se ejecutará. Entendida la dependencia $S_3 \delta S_1$ como una relación de orden de ejecución que indica que la iteración i de S_1 no puede ejecutarse hasta que la iteración $i-3$ de S_3 haya sido ejecutada, se tiene que, después de ejecutar la iteración $i-3$ de S_2 , que es la que decide la ejecución de S_3 , se puede iniciar la ejecución de la iteración i de S_1 . Esto implica que el subgrafo de dependencias, en caso de no cumplirse nunca la condición evaluada en S_2 , queda modificado tal como se indica en la figura 3.7(c). En este subgrafo denotamos con un arco discontinuo las relaciones de dependencia causadas por los efectos laterales de dependencias en otras ramas de ejecución. El paralelismo en este caso siempre es 3 tanto si se ejecuta o no la rama condicional.

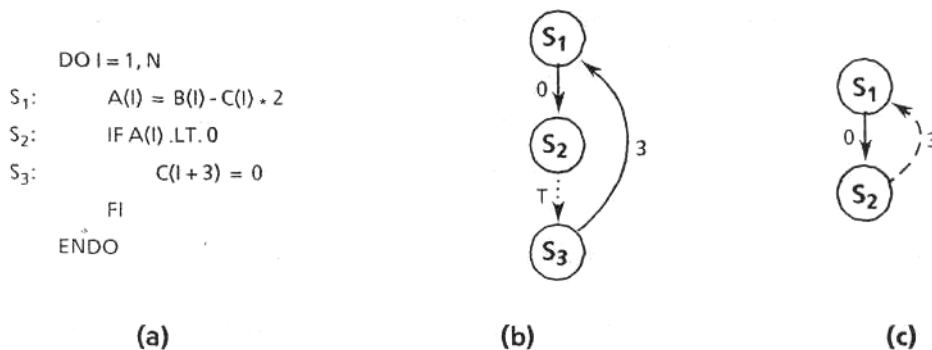


Figura 3.7: Bucle con sentencia estructurada condicional.

Construcción del subgrafo de dependencias

En general, dada una determinada rama de ejecución V_i , y siendo S_c la sentencia que decide la ejecución de las ramas, cualquier arco del grafo d_{lm} entre el nodo S_i incluido en otra rama de ejecución V_j y el nodo S_m incluido en V_i equivale a:

una dependencia $S_c \delta S_m$ siempre y cuando S_m pertenezca a ambas ramas de ejecución, es decir, $S_m \in V_i \cap V_j$ tal como muestra la figura 3.8(a). La distancia asociada a dicho arco es $d_{cm} = d_{lm}$. Si no es así, es decir, el nodo S_m está dentro de la sentencia estructurada condicional, no es necesario considerarlo dado que originalmente existe una dependencia de control $S_c \delta^c S_m$ con

distancia asociada $d_{cm} = 0$ más restrictiva que la anterior. La figura 3.8(b) ilustra esta idea.

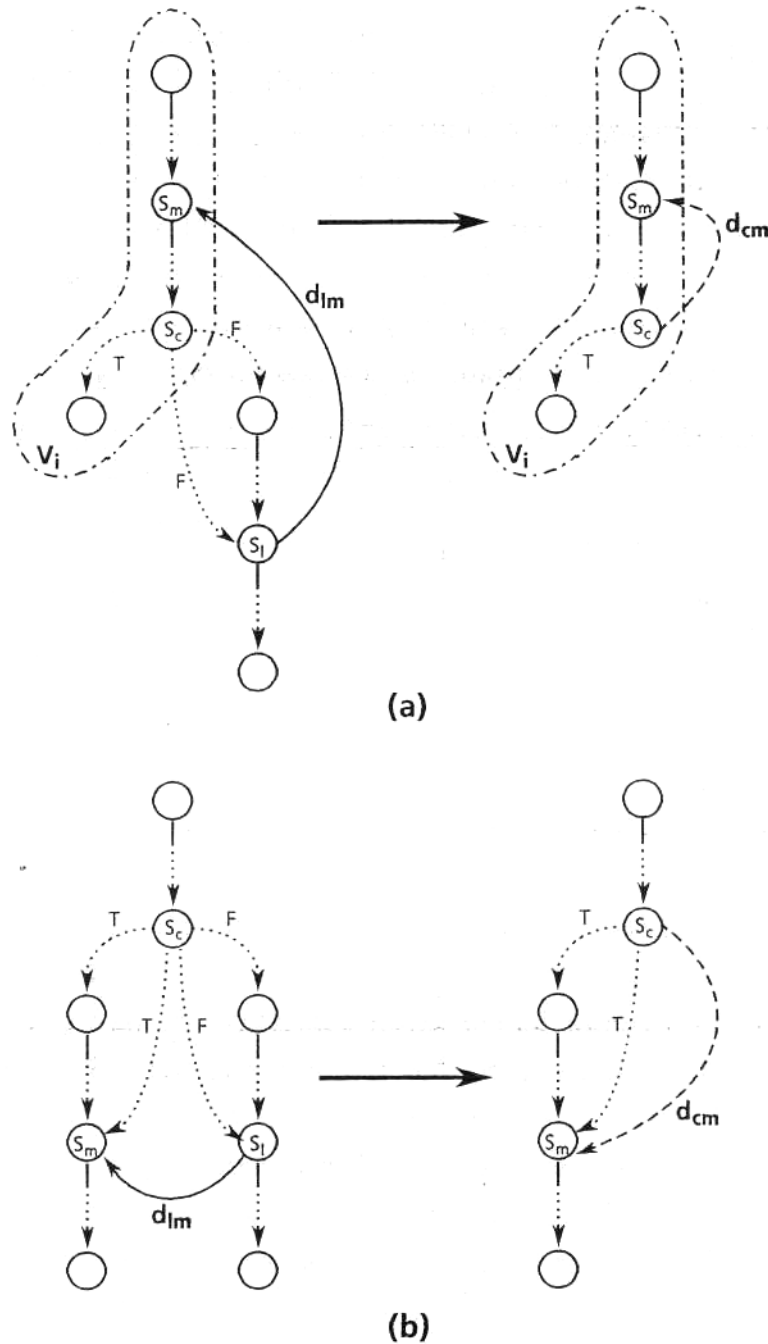


Figura 3.8: Efectos laterales de arcos de dependencias.

A partir del grafo de dependencias $G(V, E)$, el subgrafo de dependencias asociado a cada posible rama de ejecución $G_s(V_s, E_s)$ queda definido por el conjunto de sentencias V_s que se ejecutan en dicha rama y por las relaciones de dependencia E_s directas o debidas a efectos laterales de la no ejecución del resto de ramas posibles

$$E_s = \{d_{ij} \mid S_i, S_j \in V_s \wedge S_i \delta S_j\} + \{d_{cm} \mid ((S_c, S_m \in V_s) \wedge (\exists d_{lm} \mid S_c \delta^c S_l, S_l \notin V_s))\}.$$

A continuación se muestran algunos ejemplos que ilustran esta idea. La figura 3.9(a) muestra el grafo de dependencias asociado a un bucle que incluye una sentencia estructurada tipo IF-THEN-ELSE sin dependencias entre sentencias incluidas en ramas de ejecución distintas. Las figuras 3.9(b) y 3.9(c) muestran los subgrafos de dependencia obtenidos para las dos posibles ramas de ejecución. Observar en este caso que el arco d_{31} o el asociado a su efecto lateral d_{21} determinan el paralelismo de ambas ramas de ejecución. En este caso la cota superior del paralelismo es 6 y se alcanza cuando se obtiene siempre 'falso' en la condición evaluada en la sentencia S_2 .

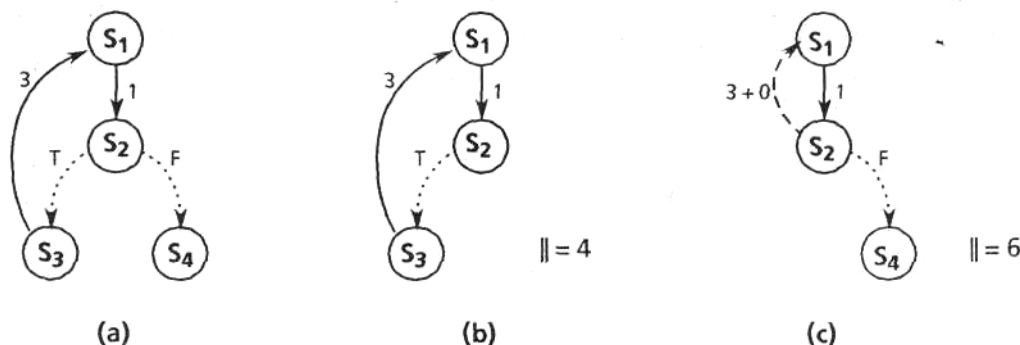


Figura 3.9: Bucle con IF-THEN-ELSE sin dependencias entre ramas condicionales.

La figura 3.10(a) muestra otro grafo de dependencias que incluye una sentencia condicional del tipo IF-THEN-ELSE con dependencias entre sentencias incluidas en ramas de ejecución distintas. En las figura 3.10(b) y 3.10(c) se muestran los subgrafos de dependencia y el paralelismo evaluado. Observar que el arco d_{41} o el asociado a su efecto lateral d_{21} con distancia asociada 1 determinan el paralelismo de ambas ramas de ejecución. El arco de dependencia d_{34} no ejerce influencia en ninguna de las dos ramas. La cota superior del paralelismo es en este caso 3 y se alcanza cuando se obtiene siempre 'cierto' en la condición establecida en la sentencia S_2 .

La figura 3.11(a) muestra el grafo de dependencias para un bucle en el que aparecen varias sentencias condicionales anidadas. Las figuras 3.11(b)-(d) muestran los subgrafos de dependencia asociados a cada posible rama de ejecución y el paralelismo evaluado. En este ejemplo se tiene en cuenta la propiedad transitiva de las dependencias de control a la hora de obtener las dependencias debidas a efectos laterales. Así por ejemplo, en el subgrafo de la figura 3.11(b) se observa el arco d_{21} debido al efecto lateral de

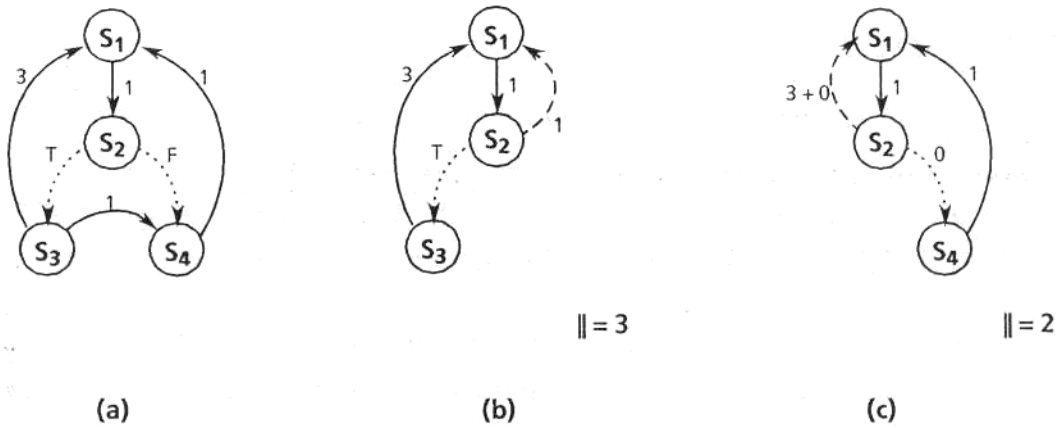


Figura 3.10: Bucle con IF-THEN-ELSE con dependencias entre ramas condicionales.

$S_2 \delta^c S_5 \delta^c S_7 \delta S_1$. En este caso el arco d_{41} determina el paralelismo de la rama $\{S_1, S_2, S_3, S_4\}$ mientras que el arco d_{71} o su efecto lateral asociado d_{51} determinan el paralelismo de las otras dos ramas posibles. En este caso, la cota superior del paralelismo es 5.

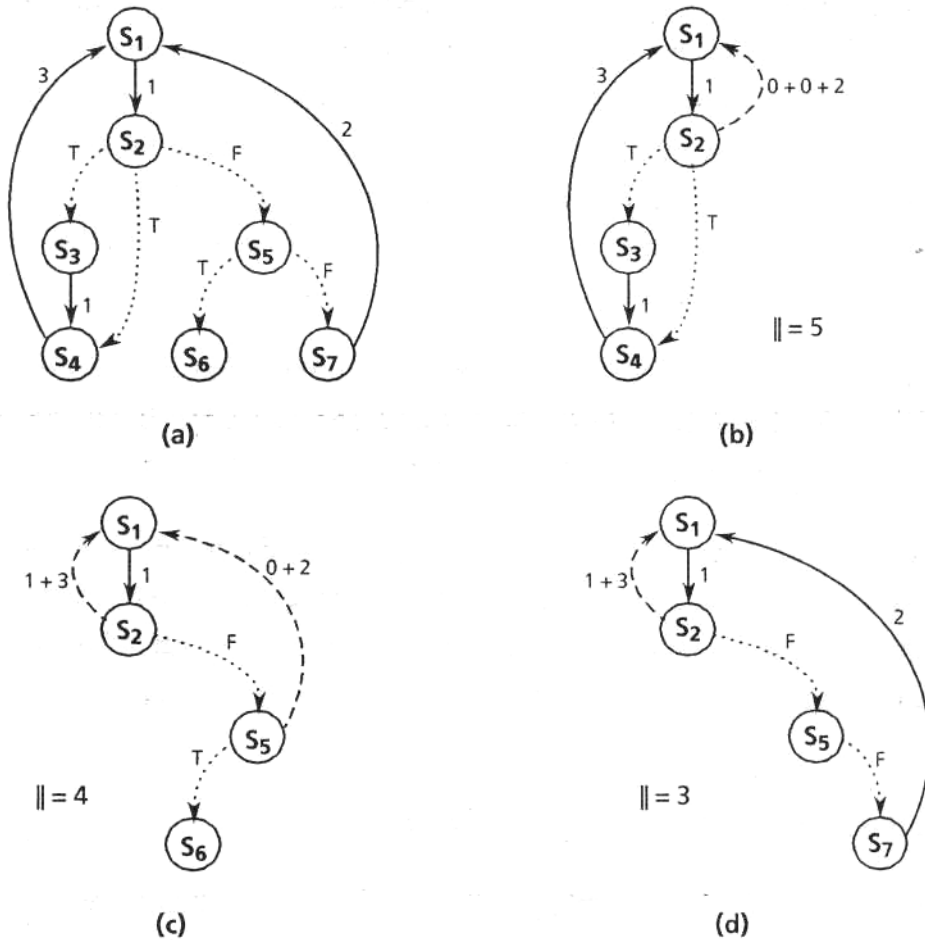


Figura 3.11: Bucle con sentencias condicionales anidadas.

3.4.1.2 Cota inferior del paralelismo

La cota inferior del paralelismo viene dada por el mínimo de los paralelismos asociados a cada una de las recurrencias que aparecen en el grafo de dependencias. Estas recurrencias pueden incluir nodos pertenecientes a una única rama de ejecución o nodos pertenecientes a varias ramas de ejecución.

Para cada recurrencia, el paralelismo mínimo se obtiene buscando el patrón o secuencia de ejecución de las distintas ramas para el cual dicha recurrencia limita al máximo.

Secuencia de ejecución de las ramas determinante del paralelismo

Si se considera el grafo de dependencias de la figura 3.12(a) y los posibles subgrafos para las dos ramas de ejecución (figuras 3.12(b)-(c)), se observa que, en principio, el paralelismo del bucle sería $\parallel = N$, siendo N el número de iteraciones del bucle a realizar. Sin embargo, si se considera la ejecución desarrollada de la figura 3.12(d), en la cual se ha fijado un patrón cíclico de ejecución de ambas ramas condicionales, se puede observar que existe un camino a través de los nodos 3 y 4 que determina el tiempo de ejecución del bucle. Si se aplican a este camino las expresiones del apartado 3.1.3, se obtiene un paralelismo por sentencia $\parallel_s = 3/2$ y por lo tanto, un paralelismo del bucle de $\parallel = 3 \cdot 3/2 = 4.5$, inferior al evaluado para las ramas de ejecución.

La secuencia de ejecución de las ramas que determina el paralelismo de una recurrencia viene dado por las distancias asociadas a los arcos que constituyen la recurrencia.

En el grafo de la figura 3.12, si se considera $V_1 = \{S_1, S_2, S_3\}$ y $V_2 = \{S_1, S_2, S_4\}$ como las dos posibles ramas de ejecución, el patrón viene dado por la ejecución de V_1 en la iteración i , V_2 en la iteración $i + 2$, V_1 en la iteración $i + 3$ y así sucesivamente, siendo independiente la rama ejecutada en la iteración

Influencia del número de sentencias de las ramas

Si se supone el grafo de dependencias de la figura 3.13(a) y se consideran $V_1 = \{S_1, S_2, S_3\}$ y $V_2 = \{S_1, S_2, S_4, S_5\}$ como las dos posibles ramas de ejecución, es fácil de ver que el paralelismo máximo del bucle es 5 y que se alcanza cuando la condición evaluada en S_2 da siempre 'falso', es decir, cuando

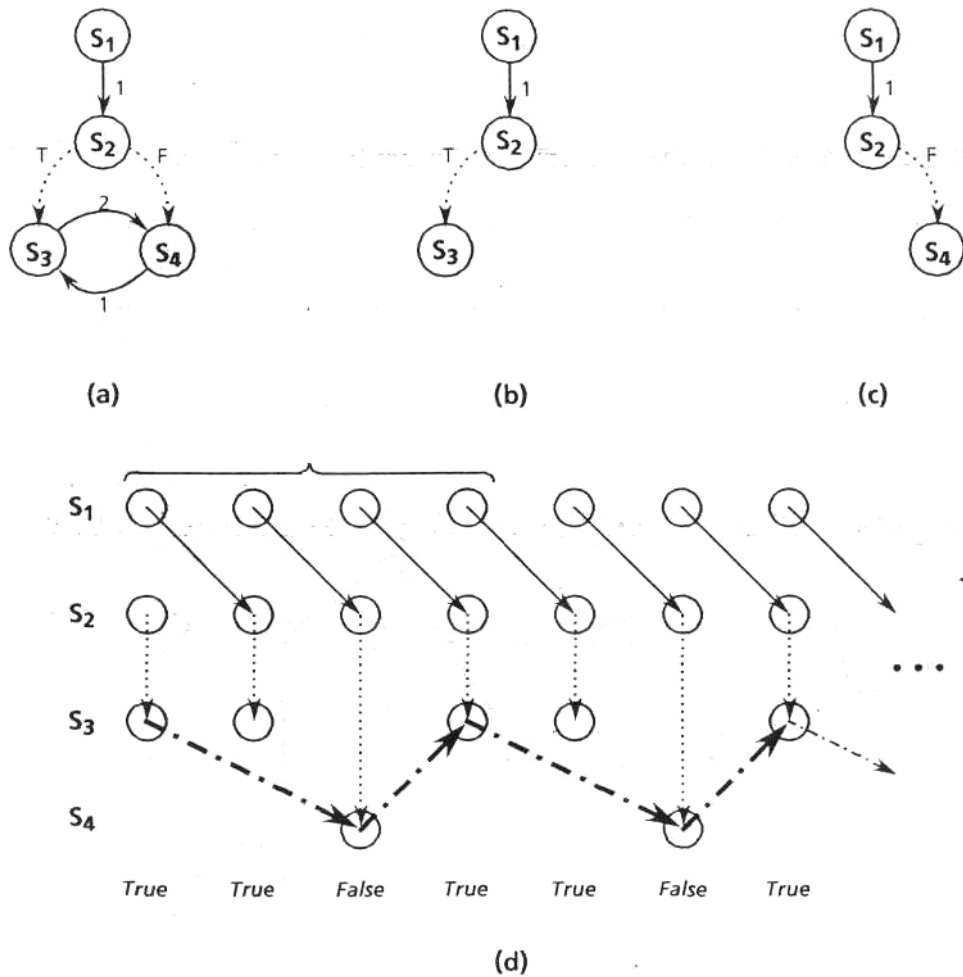


Figura 3.12: Limitación del paralelismo por recurrencias entre ramas de ejecución.

se ejecuta siempre la rama V_2 . En el grafo se observan 6 recurrencias: $R_1 = \{d_{12}, d_{23}, d_{31}\}$, $R_2 = \{d_{12}, d_{24}, d_{45}, d_{51}\}$, $R_3 = \{d_{12}, d_{25}, d_{51}\}$, $R_4 = \{d_{12}, d_{23}, d_{34}, d_{45}, d_{51}\}$, $R_5 = \{d_{12}, d_{24}, d_{45}, d_{53}, d_{31}\}$ y $R_6 = \{d_{34}, d_{45}, d_{53}\}$. Para cada una de ellas se puede buscar cual es el patrón de ejecución para el cual la recurrencia limita al máximo. Así por ejemplo, el patrón de ejecución de las ramas condicionales para el cual R_2 limita al máximo viene dado por la ejecución de V_2 en las iteraciones i e $i+1$, la ejecución de la rama con cardinalidad menor en la que se encuentra S_1 , es decir V_1 en la iteración $i+5$, de V_2 en la iteración $i + 6$ y así sucesivamente. Observar que en las iteraciones $i+2, i+3$ e $i + 4$ es indiferente la rama que se ejecuta. Para este patrón se obtiene la ejecución desarrollada de la figura 3.13(d), en la que sólo se muestran las dependencias involucradas en la recurrencia considerada.

Para aplicar a este camino las expresiones del apartado 3.1.3 que permiten evaluar el paralelismo de una recurrencia, se debe considerar el número medio

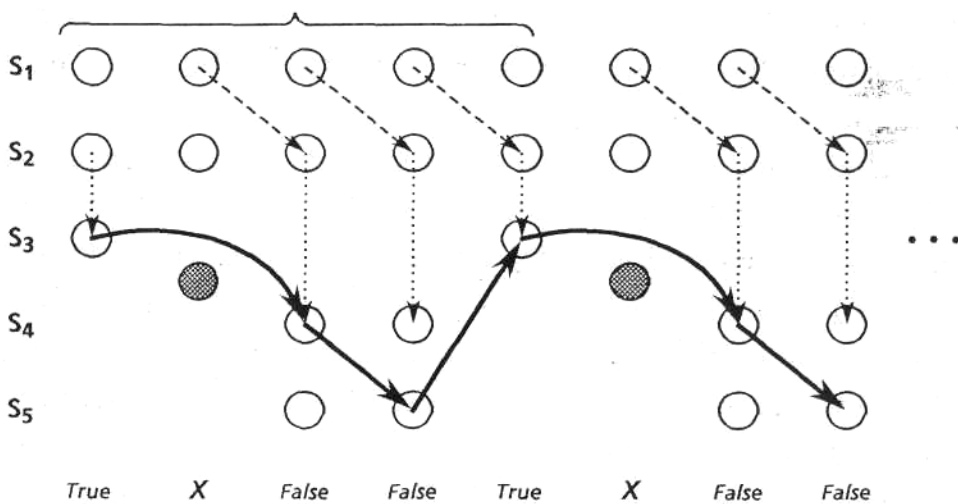
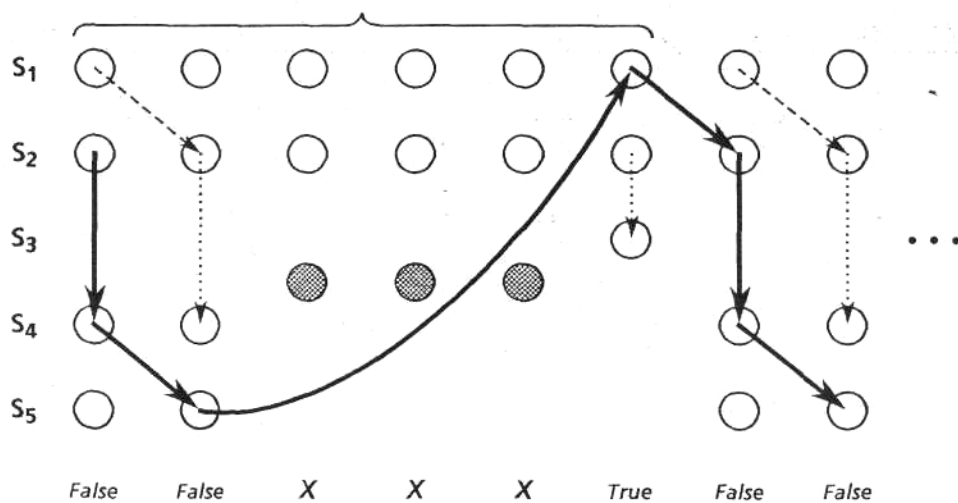
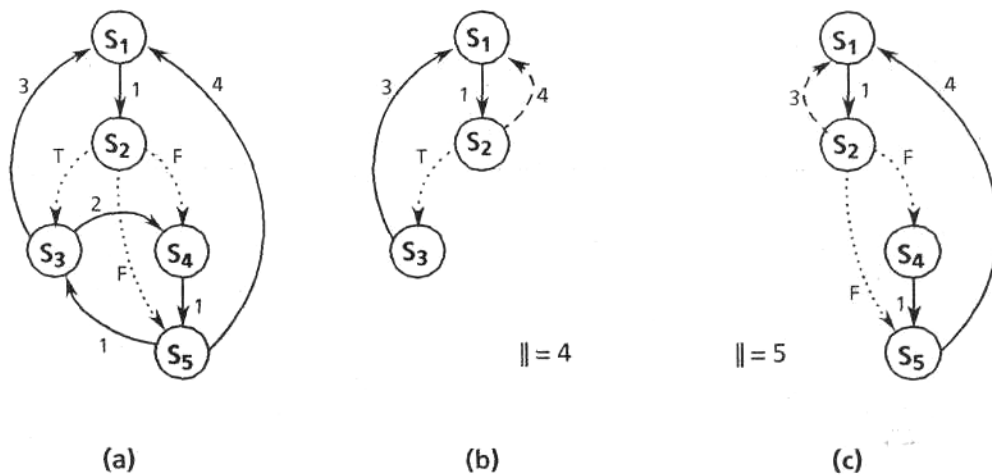


Figura 3.13: Evaluación del paralelismo mínimo de recurrencias en bucles condicionales.

de instanciaciones de sentencias que se van realizar durante toda la ejecución del bucle. Si las ramas de ejecución poseen el mismo número de sentencias n no aparece ningún problema adicional. Si no es así, hay que considerar el número de instanciaciones de sentencias ejecutadas, y en el caso de buscar una cota inferior del paralelismo, el menor número de estas que se pueden ejecutar.

Para la ejecución desarrollada de la figura 3.13(d) se tiene que en cada paso por la recurrencia se avanzan 6 iteraciones y por lo tanto, durante la ejecución del bucle se van a realizar $N/6$ pasadas por esta recurrencia, siendo N el número de iteraciones del bucle a realizar. Dado que el tiempo de ejecutar una paso por la recurrencia es de 4 ciclos, se tiene que el tiempo de ejecutar esta cadena de dependencias es $(4/6 \cdot N)$. Por lo tanto, el paralelismo viene determinado por el número total de operaciones a realizar dividido por el tiempo de ejecutar esta cadena de dependencias. Sin embargo, el número total de operaciones depende de las ramas que se ejecuten en las iteraciones $i+2$, $i+3$ e $i+4$ y que en el peor de los casos, corresponde a la ejecución de la rama V_1 , pues es la que tiene asociada una cardinalidad menor. En este caso, el número medio de operaciones que se realizan es $10/3$ y por lo tanto el total de operaciones sería $(10/3 \cdot N)$. El paralelismo del bucle es, en el peor de los casos es $\parallel = 10/3 - 6/4 = 5$.

Si sobre el mismo grafo de dependencias de la figura 3.13(a) consideramos la recurrencia R_6 , se tiene que el patrón que limita el paralelismo viene dado por la ejecución de V_1 en la iteración i , de V_2 en la iteración $i+2$, de V_2 en la iteración $i+3$, de V_1 en la iteración $i+4$ y así sucesivamente. Para este patrón de ejecución se obtiene la ejecución desarrollada de la figura 3.13(e), en la que sólo se muestran las dependencias involucradas en la recurrencia que cubre S_3 , S_4 y S_5 . Observar, que la rama que se ejecuta en la iteración $i+1$ es indiferente. Aplicando el mismo método que en la recurrencia anterior se obtiene que el paralelismo mínimo de esta recurrencia es $14/3$.

Paralelismo de una recurrencia

En general, el paralelismo \parallel de una recurrencia R que incluye nodos pertenecientes a diferentes ramas de ejecución viene dado por

$$\parallel(R) = \bar{n}(R) \cdot \frac{w(R)}{|R|}$$

siendo $\bar{n}(R)$ el número medio de sentencias ejecutado en cada iteración del bucle para el patrón de ejecución que determina dicha recurrencia. Este

número medio viene determinado (a) por las veces que se ejecuta cada rama multiplicado por la cardinalidad de la misma y (b) por el número de iteraciones en que es indiferente que rama se ejecute multiplicado por la mínima de las cardinalidades de las ramas involucradas en la recurrencia.

Se define r_i como la rama de ejecución con cardinalidad menor en la que está incluida la sentencia S_j , es decir,

$$r_i = V_k \mid ((S_i \in V_k) \wedge (|V_k| \leq |V_l|, \forall V_l, S_i \in V_l)).$$

Dado que la ejecución de una sentencia de la recurrencia implica la ejecución de la rama que la contiene, se tiene que el número total de sentencias ejecutadas en un paso por la recurrencia es

$$n_1 = \sum_{d_{ij} \in R \mid d_{ij} \neq 0} |r_i|.$$

Por otro lado se tiene que cada arco d_{ij} de la recurrencia R con distancia $d_{ij} > 1$ genera $d_{ij} - 1$ iteraciones en las que es indiferente que rama se ejecuta. Por lo tanto, a fin de obtener una cota mínima del paralelismo se deberá considerar la rama que menos sentencias ejecute. De esta manera, se tiene que el número total de sentencias ejecutadas en las iteraciones indiferentes es

$$n_2 = \min_{V_k} |V_k| \cdot \sum_{d_{ij} \in R} \max(0, d_{ij} - 1).$$

Por lo tanto, el número medio de sentencias \bar{n} ejecutadas en una pasada viene dado por

$$\bar{n} = \frac{n_1 + n_2}{w(R)}.$$

Cota inferior del paralelismo

La cota inferior del paralelismo viene dada por el mínimo de los paralelismo de cada una de las recurrencias que incluye el grafo de dependencias

$$\parallel = \min_{R \in B} \left(\bar{n}(R) \cdot \frac{w(R)}{|R|} \right)$$

3.4.2 Modelo probabilístico del paralelismo

En este apartado se considera la influencia de la probabilidad de las ramas de ejecución en el paralelismo máximo del bucle. Esta influencia se muestra en gráficos obtenidos por simulación. En estas simulaciones se considera a como la probabilidad de que dé 'cierta' (true) la condición evaluada en la sentencia que determina la ejecución condicional. Como hipótesis de trabajo se considera que la probabilidad de ejecución de las ramas es independiente de los cálculos realizados en el bucle. Por otro lado dichos resultados se contrastan con cotas analíticas del paralelismo máximo en función de la probabilidad obtenidas directamente a partir del grafo de dependencias.

Caso 1: una única recurrencia que limita el paralelismo

La figura 3.14(c) muestra los resultados obtenidos para un bucle con grafo de dependencias de la figura 3.14(a). Este grafo de dependencias se caracteriza por poseer una única recurrencia que limita el paralelismo de las posibles ramas de ejecución. La expresión analítica que determina la variación del paralelismo máximo en función de la probabilidad a es función de (a) el número de operaciones a realizar y (b) el tiempo necesario para ejecutarlas siguiendo la cadena de dependencias más larga. En la figura 3.14(b) se observan las dos posibles cadenas de dependencias que se generan según el resultado de la condición evaluada. Puede observarse que si la condición evaluada da 'cierto' el tiempo requerido para ejecutar un paso por la recurrencia es 3 mientras que si da 'falso' el tiempo requerido es 2. Por lo tanto, si a es la probabilidad de obtener 'cierto' en la condición evaluada, se tiene que el tiempo requerido para ejecutar un paso por la recurrencia es

$$a \cdot 3 + (1 - a) \cdot 2.$$

Dado que el número de pasos por la recurrencia durante la ejecución completa del bucle viene determinado por el número total de iteraciones a realizar N y por el peso de la misma, se tiene que el paralelismo es

$$\parallel = \frac{3 \cdot N}{(a \cdot 3 + (1 - a) \cdot 2) \cdot \frac{N}{4}} = \frac{12}{2 + a}$$

que en la gráfica de la figura 3.14(c) se representa en trazo discontinuo. En este caso, la expresión del paralelismo obtenido es exacta dado que todas las cadenas de dependencia que aparecen en la ejecución desarrollada son independientes.

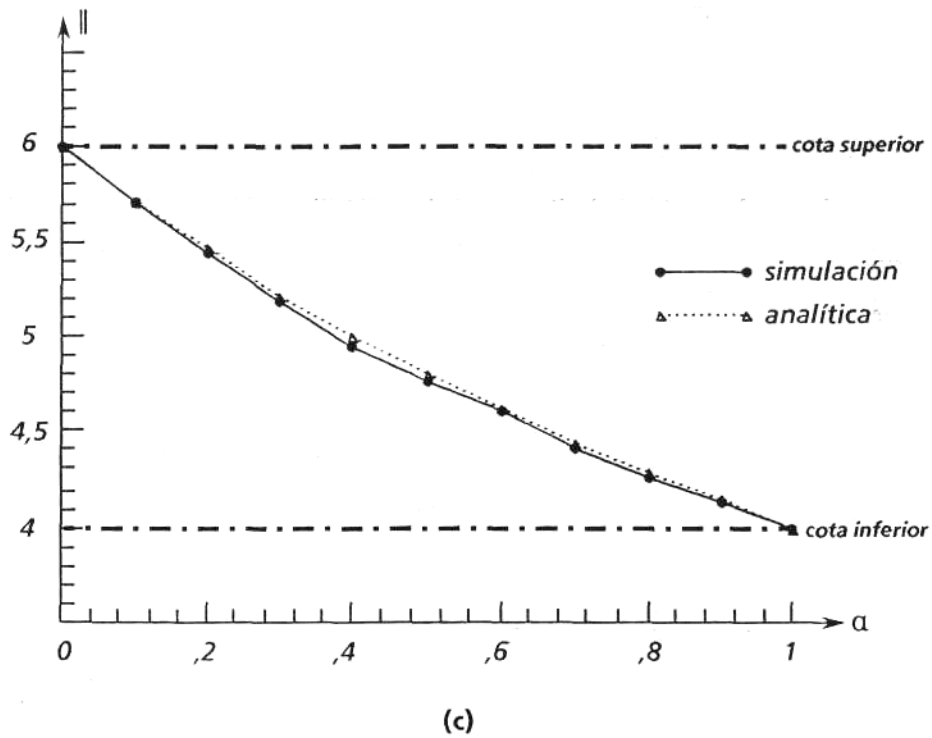
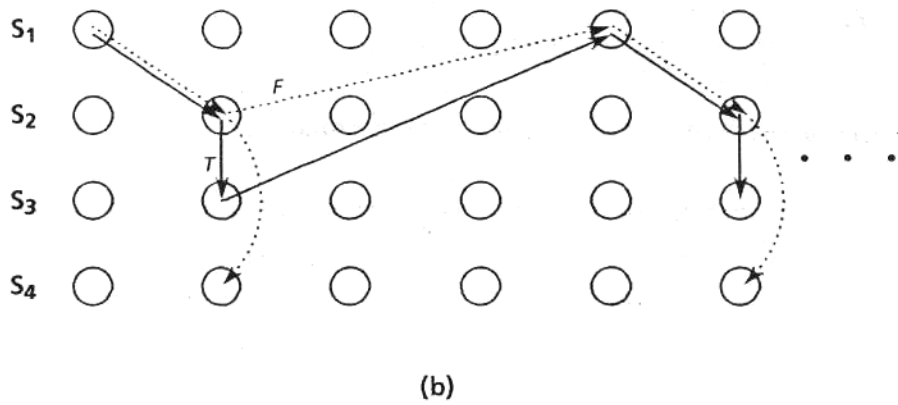
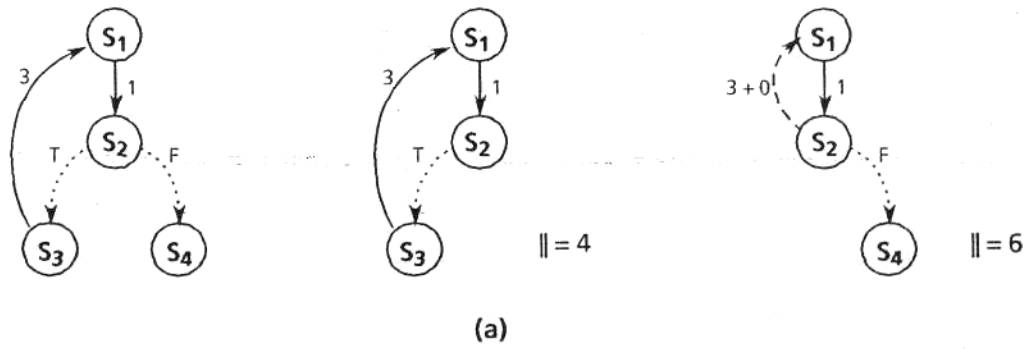


Figura 3.14: Variación del número de ciclos en las cadenas de dependencias.

Caso 2: influencia de otras dependencias

La figura 3.15(c) muestra los resultados obtenidos para el grafo de dependencias de la figura 3.15(a). Este grafo se caracteriza por tener varias recurrencias aunque el paralelismo de ambas ramas de ejecución viene determinado por el mismo arco de dependencia d_{42} , tal como muestran los subgrafos asociados. En éste caso la expresión analítica se obtiene de forma similar al caso anterior y es

$$\| \leq \frac{3 \cdot N}{(\alpha \cdot 2 + (1 - \alpha) \cdot 3) \cdot \frac{N}{3}} = \frac{9}{3 - \alpha}$$

Sin embargo, en este caso, la expresión no es exacta dado que no se considera el efecto del arco d_{31} que puede introducir retardos en la ejecución de las cadenas de dependencia. Esta expresión analítica es una cota superior del paralelismo máximo del bucle.

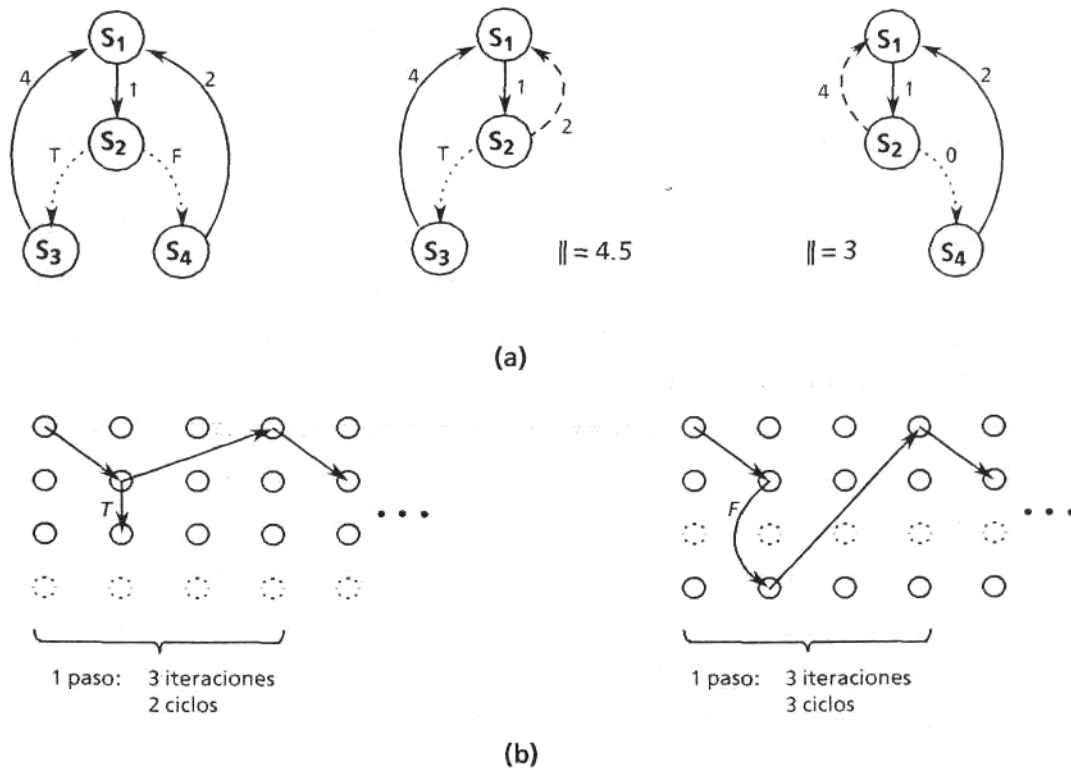


Figura 3.15: Influencia de otras dependencias que no determinan las cadenas de dependencia.

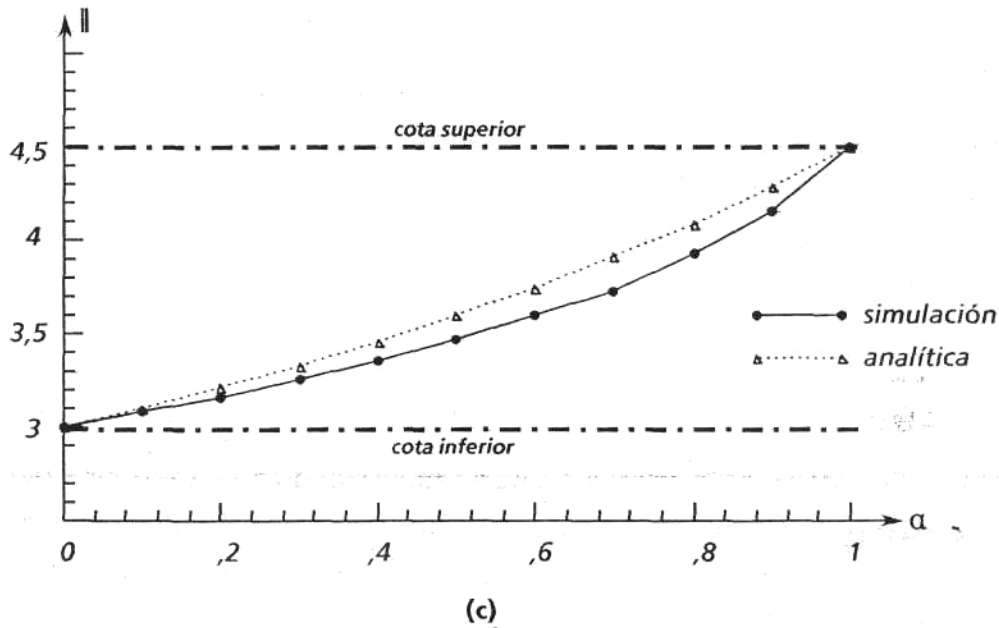


Figura 3.15 (cont): Influencia de otras dependencias que no determinan las cadenas de dependencia.

Caso 3: varias recurrencias que limitan el paralelismo

A continuación se considera el caso de que el paralelismo de las ramas de ejecución venga limitado por recurrencias distintas. La figura 3.16(c) muestra los resultados obtenidos para el grafo de dependencias de la figura 3.16(a). Este grafo se caracteriza por tener dos recurrencias que limitan el paralelismo de las dos posibles ramas de ejecución, tal y como muestran los subgrafos asociados. A fin de obtener la expresión analítica del paralelismo en función de la probabilidad α , es necesario obtener el tiempo de ejecución de la cadena de dependencias más larga. Tal como muestra la figura 3.16(b), se requieren 3 ciclos para ejecutar un paso por cualquiera de las recurrencias que limitan el paralelismo de las ramas. Sin embargo, y a diferencia de los ejemplos anteriores, si la condición evaluada da 'cierto' se avanzan 5 iteraciones en un paso por la recurrencia respectiva, y si da 'falso' se avanzan 4 iteraciones en un paso por la recurrencia respectiva. Por lo tanto, el número de pasos por las recurrencias que se realiza durante la ejecución de las N iteraciones del bucle viene dado por

$$\alpha \cdot \frac{N}{5} + (1 - \alpha) \cdot \frac{N}{4}$$

Por lo tanto, el paralelismo del bucle viene determinado por

$$\| \leq \frac{3 \cdot N}{3 \cdot \left(\alpha \cdot \frac{N}{5} + (1-\alpha) \cdot \frac{N}{4} \right)} = \frac{20}{5-\alpha} .$$

En este caso, la expresión analítica obtenida es también una cota superior del paralelismo real ya que existen dependencias entre las cadenas consideradas.

Caso 4: recurrencias internas a una rama de ejecución

A continuación se considera el efecto de recurrencias que limitan el paralelismo de una rama de ejecución y que no engloban a todas las sentencias de dicha rama. La figura 3.17(a) muestra un grafo de dependencias asociado a un bucle tipo IF-THEN en el que la recurrencia $\{d_{33}\}$ limita el paralelismo de la rama de ejecución asociada al THEN, tal y como muestran los subgrafos de la figura. La figura 3.17(c) muestra con trazo continuo los resultados obtenidos mediante simulación para este bucle. La obtención de la expresión analítica para la cota superior del paralelismo se complica en estos casos. En primer lugar, el número de operaciones del bucle a realizar depende de la probabilidad a de las ramas de ejecución, y viene dado por

$$(2 + a) \cdot N .$$

En segundo lugar, en la figura 3.17(b) puede observarse que la obtención de 'falso' en la condición evaluada en S_2 permite un avance de 5 iteraciones en 2 ciclos. Sin embargo, el avance y número de ciclos tras la obtención de un 'cierto' depende del número de 'ciertos' i consecutivos que le sigan. Por cada 'cierto' que resulte, se avanza 1 iteración más y se tarda un ciclo más sobre las 5 iteraciones que se avanzan y 3 ciclos requeridos en el caso de producirse un único 'cierto'. Observar que un paso por la recurrencia se cierra cuando se obtiene un 'falso' que permite abandonar la recurrencia $\{d_{33}\}$. Por lo tanto, el tiempo requerido para ejecutar una cadena de dependencias viene dado por

$$(1-\alpha) \cdot 2 \cdot \frac{N}{5} + \sum_{i=0}^{\infty} (3+i) \cdot \frac{N}{(5+i)} \cdot \alpha^{i+1} \cdot (1-\alpha) .$$

La cota superior del paralelismo viene dada por

$$\| \leq \frac{(2 + a) \cdot N}{(1-\alpha) \cdot 2 \cdot \frac{N}{5} + \sum_{i=0}^{\infty} (3+i) \cdot \frac{N}{(5+i)} \cdot \alpha^{i+1} \cdot (1-\alpha)} = \frac{(2 + a)}{(1-\alpha) \cdot \frac{2}{5} + \sum_{i=0}^{\infty} \frac{3+i}{5+i} \cdot \alpha^{i+1} \cdot (1-\alpha)} .$$

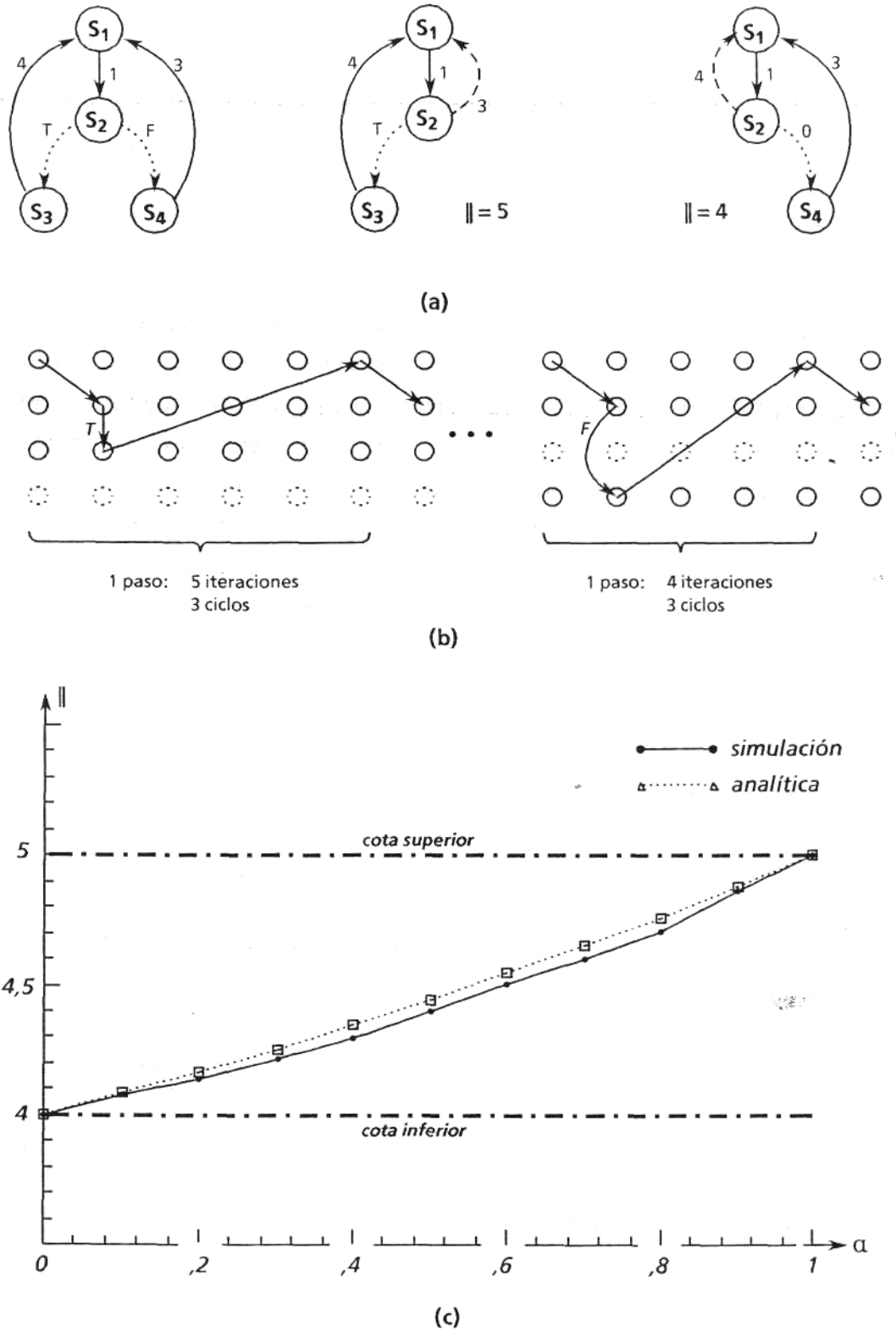


Figura 3.16: Variación del número de iteraciones cubiertas por las cadenas de dependencias.

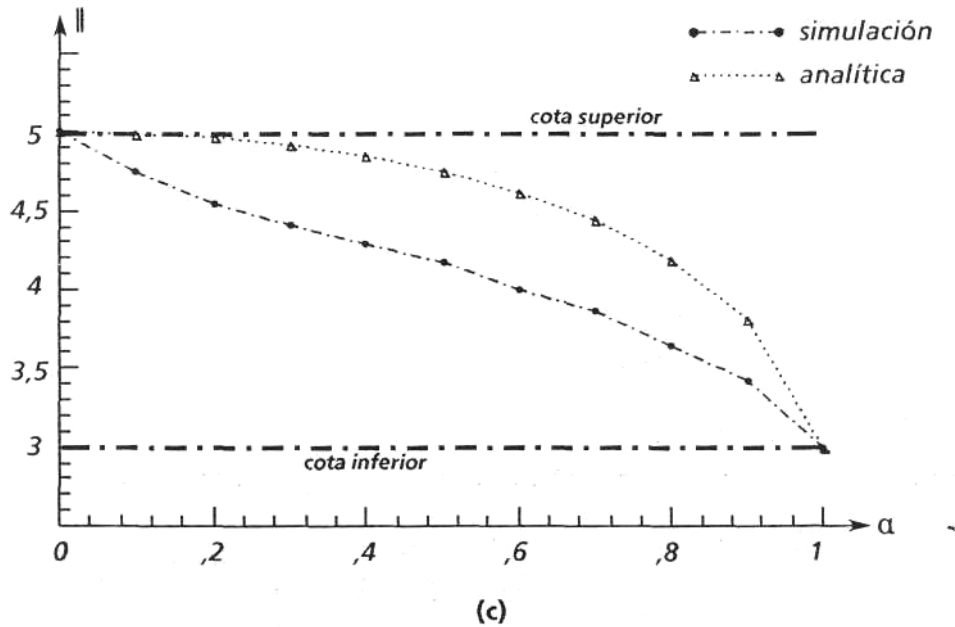


Figura 3.17 (cont): Efecto de recurrencias internas a una rama de ejecución.

diferentes ramas de ejecución. En la figura 3.18(b) se observan las cuatro cadenas de dependencia que pueden aparecer en función de los 'ciertos' y 'falsos' que aparezcan. Por ejemplo, para la primera cadena mostrada se observa que cada paso por la recurrencia $\{d_{34}, d_{43}\}$ se produce con una probabilidad de

$$(1 - \alpha) \cdot \alpha$$

y que provoca un avance adicional de 2 iteraciones sobre las 5 que en principio se avanzarían. Además, cada paso por la recurrencia requiere 2 ciclos adicionales para ejecutarse sobre los 3 originales. Esta cadena se inicia con la aparición de un 'cierto', seguido de i parejas 'falso-cierto' y finaliza con la aparición de un 'cierto', que obliga a abandonar la recurrencia $\{d_{34}, d_{43}\}$. Por lo tanto, el tiempo requerido para ejecutar esta cadena es

$$\sum_{i=0}^{\infty} (3+i) \cdot \frac{N}{(5+i)} \cdot \alpha^{i+2} \cdot (1-\alpha)^i$$

Si se aplica la misma idea para todas las cadenas mostradas en la figura 3.18(b) se obtiene como expresión del paralelismo

$$\| \leq \frac{3}{\sum_{i=0}^{\infty} \frac{3+2 \cdot i}{5+2 \cdot i} \cdot \alpha^{i+2} \cdot (1-\alpha)^i + \sum_{i=0}^{\infty} \frac{4+2 \cdot i}{5+2 \cdot i} \cdot \alpha^{i+1} \cdot (1-\alpha)^{i+2} + \sum_{i=0}^{\infty} \frac{3+2 \cdot i}{4+2 \cdot i} \cdot \alpha^i \cdot (1-\alpha)^{i+2} + \sum_{i=0}^{\infty} \frac{4+2 \cdot i}{6+2 \cdot i} \cdot \alpha^{i+2} \cdot (1-\alpha)^{i+1}}$$

En la figura 3.18(c) se muestra con trazo discontinuo la representación de esta cota superior analítica.

3.5 EFICIENCIA DEL PROCESO DE PARALELIZACIÓN

Como resultado importante de conocer el máximo paralelismo que podemos extraer de un bucle se tiene la posibilidad de conocer la eficiencia de los métodos de paralelización utilizados.

Se define eficiencia del proceso de paralelización como

$$\eta = \frac{P}{\|}$$

siendo P el número de procesadores requeridos por la técnica de paralelización utilizada y $\|$ el paralelismo real del bucle. Esta medida de eficiencia no es del todo correcta en cuanto a que no considera el "overhead" o sobrecarga introducida por la posible sincronización que puede requerirse entre las tareas ejecutadas por los procesadores.

- Se dice que un método de paralelización es incapaz de extraer todo el paralelismo del bucle cuando obtiene una eficiencia $\eta < 1$. Por el contrario, diremos que un método de paralelización utiliza más procesadores de los necesarios cuando $\eta > 1$.

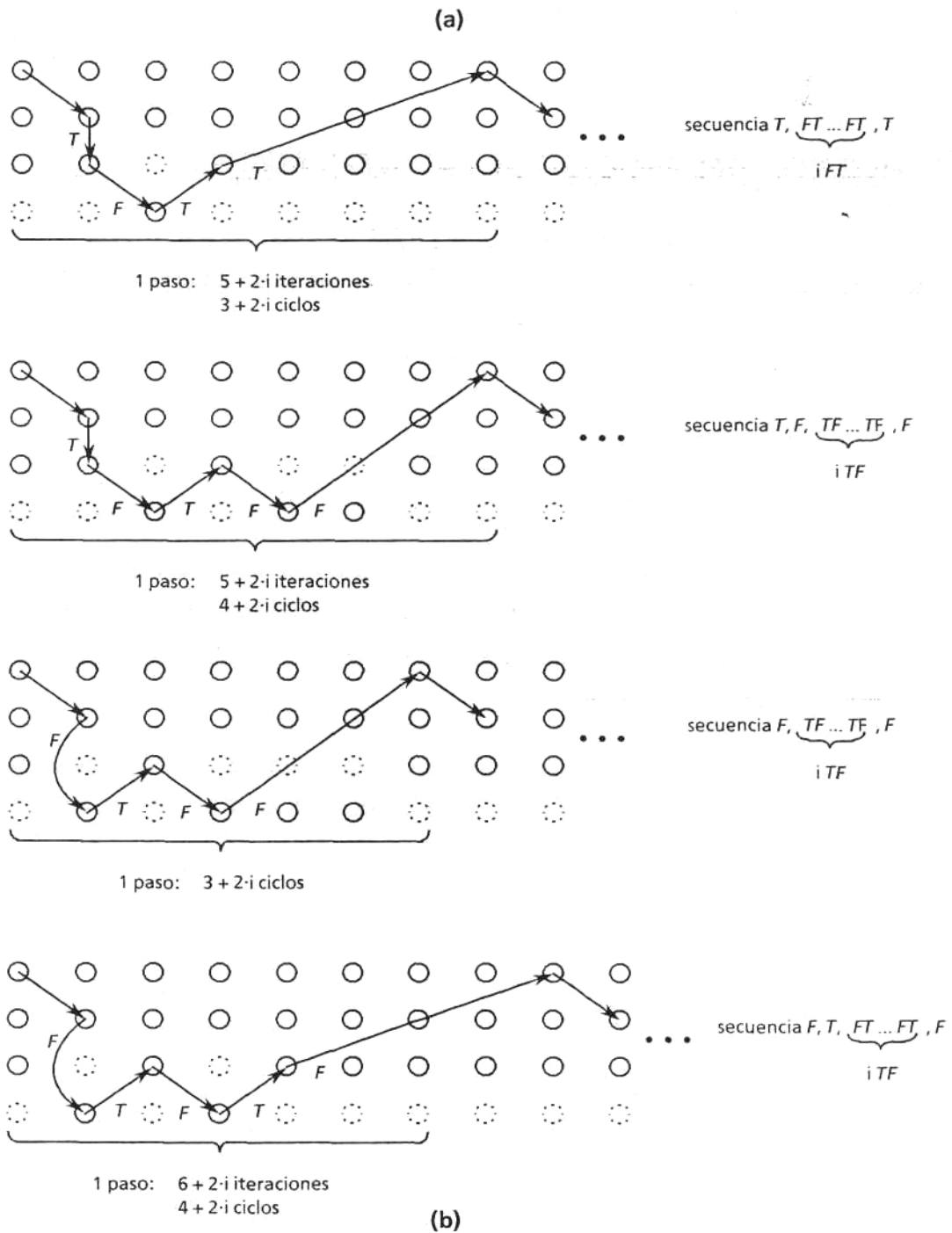
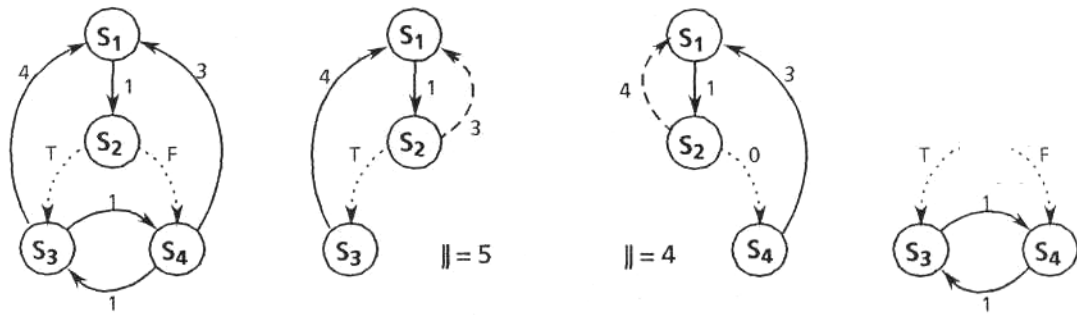


Figura 3.18: Efecto de recurrencias entre ramas de ejecución.

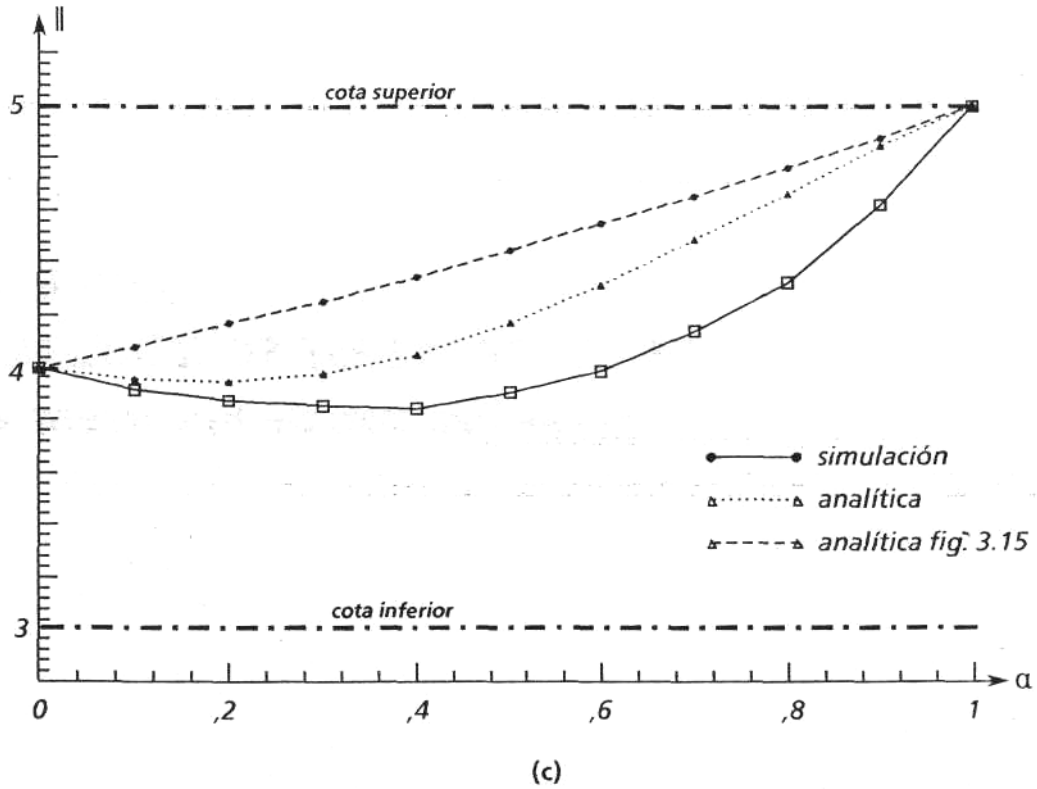


Figura 3.18 (cont): Efecto de recurrencias entre ramas de ejecución.