

## تحسين أداء شبكات الشرائح ثلاثية الأبعاد ذات

### عرض الحزمة العمودي المحدود باستخدام

#### توجيه تكيفي

م. عزيز جروج أ.د. إبراهيم الشامي

قسم هندسة التحكم الآلي والحاسوب

كلية الهندسة الميكانيكية والكهربائية - جامعة البعث

#### ملخص البحث

تعمل تقنية تكديس القوالب على توسيع التنوع المكاني للاتصالات على الرقاقة من خلال الاستفادة من التكامل (TSV) لربط الطبقات. تعاني الشبكة ثلاثية الأبعاد على الرقاقة (3D-NoC)، وهي مزيج من تقنية تكديس القوالب والبنية التحتية للاتصالات المنتظمة على الرقاقة، من زيادة الكثافة الحرارية وتبديد الحرارة غير المتوازن عبر طبقات متعددة، مما يؤثر بشكل كبير على أداء وموثوقية الرقاقة. يعد TSV الطريقة الأكثر شيوعاً لتنفيذ الروابط الرأسية. ومع ذلك، توجد صعوبات على المستوى التكنولوجي لضمان إنتاج عدد مقبول من الروابط الرأسية. لذلك، غالباً ما يكون عرض النطاق الترددي للوصلات الرأسية أصغر من الروابط الأفقية، مما يشكل عنق الزجاجة للنظام بأكمله. يقدم هذا البحث خوارزمية التوجيه التكيفي التي توزع حركة المرور للأنظمة ثلاثية الأبعاد ذات النطاق الترددي المحدود في الروابط الرأسية. تبين محاكاتها أنه في بنية شبكة  $4 \times 4 \times 4$  ثلاثية الأبعاد، يمكن أن تحقق الخوارزمية المقترحة لدينا تحسناً ملحوظاً في الأداء من حيث زمن التأخير والإنتاجية مقارنة بخوارزميات التوجيه الحالية وهي قوية من حيث أن الأداء مستقر في ظل حركة مرور مختلفة الأنماط.

#### كلمات مفتاحية:

شبكات الربط على الشريحة، خوارزمية توجيه تكيفية، تكديس القوالب.

# Performance Enhancing of 3D-NOCs with Limited Vertical Bandwidth Using Adaptive Routing

**Eng.Aziz Jarrouj**

**Prof. Ibrahim El-Shami**

*Faculty of Mechanical and Electrical Engineering, Albaath University*

*Department of Auto control and Computers*

## Abstract

Die-stacking technology is expanding the space diversity of on-chip communications by leveraging through-silicon-via (TSV) integration and wafer bonding. The 3D network-on-chip (NoC), a combination of die-stacking technology and systematic on-chip communication infrastructure, suffers from increased thermal density and unbalanced heat dissipation across multistacked layers, significantly affecting chip performance and reliability. Currently, TSV is the most popular and practical way to implement vertical links. Yet, there exist difficulties at the technological level ensuring an acceptable yield number of vertical links. Therefore, the bandwidth of vertical links is often made smaller than horizontal links, which becomes a bottleneck of the whole system. This paper presents a traffic distributing adaptive routing algorithm for 3D systems with limited bandwidth in vertical links. Our simulation with synthetic traffic pattern reveals that in a 4×4×4 3D mesh network architecture, our proposed algorithm can achieve significant performance improvement in network latency and throughput compared to existing routing algorithms and is robust since the performance is stable under different traffic patterns.

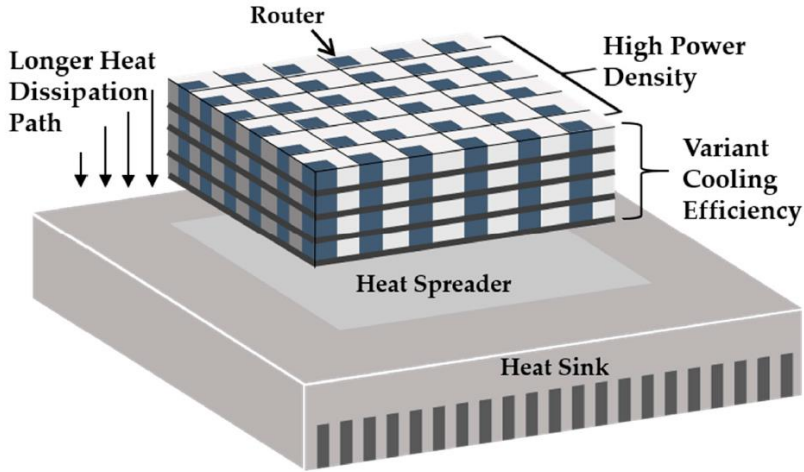
## **Keywords:**

Networks on the chip, adaptive routing algorithm, TSV.

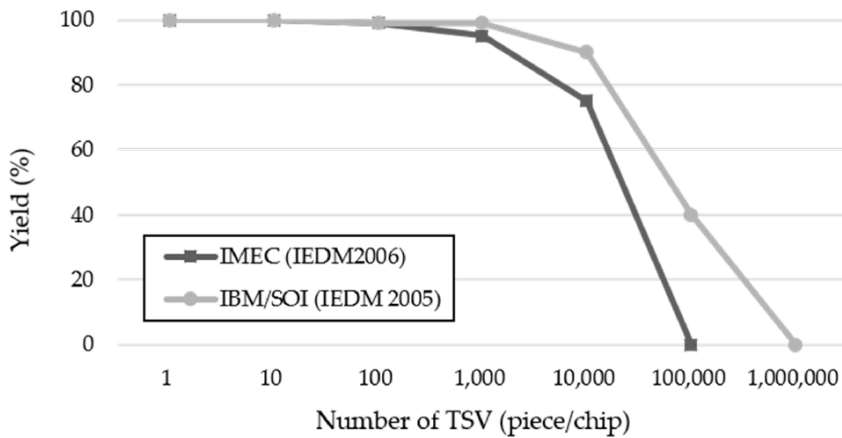
## 1. مقدمة

منذ منتصف العقد الأول من القرن الحادي والعشرين، تم استخدام رقاقة المعالجات المتعددة Chip Multi-Processor (CMP) على نطاق واسع للتغلب على القيود المتعلقة بالتوازي على مستوى التعليمات وجدران الطاقة في معالج أحادي الخيط / النواة [1]. ومع ذلك، فإن حركة المرور المتزايدة باستمرار بين عناصر المعالجة أدت إلى اختناقات في CMPs التقليدية القائمة على الناقل [2]. في البداية، تم اقتراح شبكة ثنائية الأبعاد على الرقاقة (NoC) Network-on-Chip للتخفيف من التعقيدات في شبكة التوصيل البيني على الرقاقة [3]. على الرغم من أن 2D-NoC له مزايا قابلية التوسع العالية وبنية التصنيع البسيطة، إلا أن الأداء العالي غير مضمون، بسبب التدهور السريع في زمن انتقال الحزم المرتبط بزيادة المسافات المادية مع زيادة عدد نوى المعالجة [4]. مع ظهور الدوائر المتكاملة ثلاثية الأبعاد القائمة على Through-Silicon-Via (TSV)، دخلت التغييرات الهيكلية مرحلة جديدة تتجاوز 2D-NOCs [5]. توفر بنية CMP ثلاثية الأبعاد المستندة إلى NoC والتي تستغل تكديس القوالب باستخدام تقنية TSV نطاقاً ترددياً أوسع وتأخيراً أقل في نقل الحزم ومسافات تخطيط أصغر بمتوسط مسافات أقصر من نظام NoC التقليدي ثنائي الأبعاد [6]. ومع ذلك، تؤدي الكثافة الحرارية المتزايدة في NoC ثلاثي الأبعاد في النهاية إلى التشبع في الأداء والموثوقية، مما يتطلب دوائر تبريد إضافية، مثل المشتت الحراري [7]. ينتج عن الهيكل المكسب لـ NoC ثلاثي الأبعاد مسار أطول لتبديد الحرارة وكفاءة تبريد مختلفة بين الطبقات [8]، كما هو موضح في الشكل (1). تواجه عُقد المعالج في الطبقة العليا الأبعد عن المشتت الحراري مشاكل حرارية شديدة، ويتفاقم عدم التوازن الحراري عبر الطبقات المعمارية ثلاثية الأبعاد [9]. إن TSVs المستخدمة في الهياكل المكسدة مقيدة بمناطق الترابط الأكبر، وعمليات القياس المعقدة ذات ميزات صغيرة وهناك انخفاضات حادة في العائد مع زيادة عدد TSVs [10، 11]، كما هو موضح في الشكل (2). يؤدي الانخفاض الحاد في العائد إلى الانتقال لشبكة غير منتظمة غير ثابتة Non-Stationary

والتالي تقليل مرونة توجيه الحزمة [12]. يؤدي هذا إلى انخفاض كفاءة استخدام جهاز التوجيه، والذي بدوره يزيد من مخاطر انحراف توزيع حمل المرور، مما يؤدي إلى تفاقم المشكلات الحرارية [13].



الشكل (1) الإعتبارات الحرارية في شريحة ثلاثية الأبعاد تعتمد تكديس الطبقات [8].



الشكل (2) اعتماد العائد على عدد تكاملات (TSV) لأجل عدة عمليات تصنيع [11].

في تقنية التكامل ثلاثي الأبعاد، يتم تكديس طبقات متعددة معاً باستخدام روابط عمودية. واحدة من أكثر التقنيات شيوعاً المستخدمة في الاتصال الرأسي هي TSV

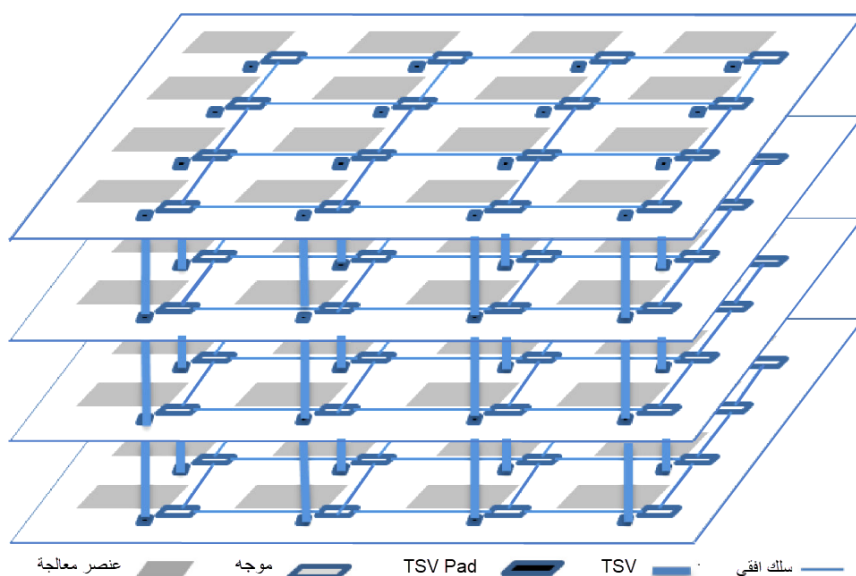
[14,15]، والتي تقوم بإنشاء اتصال بين القوالب عبر طبقات السيليكون الرقيقة. سبب شعبية TSV هو أنه يسمح بكثافة عالية، وتوافق عالي مع عملية CMOS القياسية. لسوء الحظ، تجلب TSVs أيضاً مشاكلها الخاصة التي يمكن أن تكون صعبة.

- أولاً، كما هو موضح في الشكل (3)، هناك حاجة إلى منصات هبوط TSV (TSV Pads) في كل طبقة من أجل الترابط. مع زيادة عدد عقد الشبكة في كل طبقة، يزداد أيضاً عدد TSVs، مما يؤدي إلى ارتفاع استهلاك المنطقة. على سبيل المثال،  $8 \times 8$  عقدة في كل طبقة مع 64 بت TSV عمودية لتربط العقد في الطبقات المجاورة وأبعاد محطة TSV تبلغ 10 ميكرومتر  $\times$  10 ميكرومتر. بالنسبة إلى TSV منخفضة الكثافة البالغة 50 ميكرومتر [16,17]، فإن المساحة الإجمالية لـ TSVs ستكون أكثر من 10 مم<sup>2</sup>. حتى بالنسبة إلى TSV عالية الكثافة التي تبلغ 16 ميكرومتر [16,17]، ستستهلك TSVs ما يقرب من 2.1 مم<sup>2</sup> من المساحة، والتي لا يمكن تجاهلها وتصبح مشكلة صعبة للغاية عندما يزداد حجم الشبكة.

- ثانياً، تؤدي كمية كبيرة من منصات TSV الموزعة عبر الشبكة بأكملها إلى تفاقم ازدحام التوجيه [18]، وهي مشكلة صعبة لتصميم الدوائر المتكاملة عالية السرعة.
- أخيراً، نظراً لبعض الصعوبات على المستوى التكنولوجي، فإن عمليات تصنيع TSV المتاحة حالياً لها عائد منخفض نسبياً [19] (مقارنةً بالتوصيل البيني القياسي ثنائي الأبعاد). لذلك، مع كل هذه العيوب، تميل الروابط الرأسية إلى أن يكون لها خصائص مختلفة بما في ذلك النطاق الترددي المحدود (ويرجع ذلك أساساً إلى انخفاض عدد TSVs) عند مقارنتها بالوصلات الأفقية.

لمعالجة مشكلة النطاق الترددي المحدود للوصلة الرأسية في شبكة ثلاثية الأبعاد، كانت هناك أبحاث سابقة مثل مخطط التسلسل باستخدام نطاق ساعة أعلى للإرسال العمودي [20] وهيكل ناقل هجين [21]. ومع ذلك، فهي تتطلب تصميمًا غير متزامن أو درجة عالية من التعقيد في التصميم، ما يجعل من الصعب اعتمادها. في هذا البحث، نقترح خوارزمية التوجيه التكميلي لتوزيع حركة المرور لتحسين أداء وإنتاجية الشبكة ثلاثية

الأبعاد حيث يكون عرض النطاق العمودي محدودًا. من خلال النظر في معلومات الازدحام الخاصة بالعقد المجاورة، وإعطاء أوزان مختلفة للاتجاه الرأسي والاتجاه الأفقي لحساب حركة المرور، بالإضافة إلى أخذ المسافة من العقدة الحالية إلى وجهة الحزم في الاعتبار، فإننا نعمل على تحسين زمن الانتقال والإنتاجية في الشبكة ثلاثية الأبعاد.



الشكل (3) شبكة ثلاثية الأبعاد 4x4x4 متجانسة

يعرض القسم الثاني الأعمال ذات الصلة، ويشرح القسم الثالث تفاصيل الخوارزمية المقترحة، ويوضح القسم الرابع النتائج التجريبية والتحليل. الاستنتاج معطى في القسم الخامس.

## 2. الأعمال ذات الصلة

### أ. الربط العمودي

كما ذكر أعلاه، يمثل استهلاك المنطقة والإنتاجية المنخفضة نسبيًا لـ TSVs مشكلة صعبة بالنسبة للروابط الرأسية القائمة على TSV. تم تقديم بعض الأعمال في [20]، حيث تم اقتراح مخطط تسلسلي للتوصيل البيئي الرأسي لمعالجة مثل هذه المشكلة. على سبيل المثال، يمكن أن يوفر التسلسل 4 إلى 1 لوصلات TSV أكثر من 70٪ من

مساحة منطقة TSV. ومع ذلك، سيؤدي هذا المخطط إلى انخفاض الإنتاجية خاصة عندما تكون حركة المرور غير منتظمة. علاوة على ذلك، في عملهم، يتم تطبيق مجالات ساعة مختلفة على الإرسال الأفقي والرأسي، والذي بالكاد يمكن اعتماده من قبل معظم التصميمات المتزامنة. في [17]، تم اقتراح مخطط ضغط، حيث تشترك أربع عقد شبكية في حزمة TSV واحدة. إنهم يحلون مشكلة التصميم غير المتزامن، لكن العمل يعتمد على افتراض أن أجهزة التوجيه المجاورة نادراً ما تنقل البيانات عبر قنواتها الرأسية في نفس الوقت، وهو ما قد لا يكون مقبولاً للتطبيق العملي. علاوة على ذلك، يخلق الهيكل تعقيداً إضافياً في التخطيط، مما يقلل من جدواه.

### ب. خوارزمية التوجيه ثلاثية الأبعاد

تمت دراسة خوارزمية التوجيه ثنائية الأبعاد على نطاق واسع. تؤدي خوارزميات توجيه الازدحام - الحي Congestion-oblivious مثل XY [22] و West-First [23] والعشوائية [23] إلى التوجيه دون مراعاة حالة حركة المرور للشبكة. هناك أيضاً الكثير من الخوارزميات المدركة للازدحام مثل DyXY [24] و NOP [25] و DBAR [26] و CATRA [27]، حيث يتم اختيار المسار باستخدام حالة الازدحام للشبكة. ومع ذلك، لم يتم إجراء العديد من الدراسات حول توجيه 3D NoC مقارنةً بتوجيه 2D NoC. وهناك دراسات قليلة تتناول قضية النطاق الترددي العمودي المحدود لبنية الشبكة المكسدة ثلاثية الأبعاد (3D stacked mesh).

اقترح تشاو وآخرون [28] توجيهًا تكيفيًا مدرّكًا للحرارة باستخدام توجيه تنازلي استباقي لضمان السلامة الحرارية لـ 3D NoC التي قد تعاني من الاختناق. لكن العمل يقتصر على 3D NoC المتجانسة. قامت مجموعة أخرى بحل المشكلة باستخدام هيكل هجين. اقترح رحمان وآخرون [29] استخدام ناقل وتشكيل مخطط هجين للتخفيف من الاختناق الرأسي. لقد أدخلوا أيضاً خوارزمية توجيه بين الطبقات تسمى AdaptiveZ لبنى شبكة NoC المكسدة ثلاثية الأبعاد، والتي تختار قناة عمودية وفقاً لحالة حركة المرور وتقوم بتوجيه XY في الطبقة المستهدفة. كما تداد لهذا العمل، في [21]، اقترح رحمان

وآخرون منصة مراقبة تكيفية لبنى شبكة 3D NoC المكذبة، تسمى ARBNET لإجراء مراقبة حركة المرور، والإدارة الحرارية، والتسامح مع الأخطاء. لقد أدخلوا أيضاً خوارزمية AdaptiveXYZ، وهي نسخة معدلة من خوارزمية AdaptiveZ. بدلاً من تنفيذ توجيه XY في الطبقة المستهدفة، فإنه يؤدي إلى توجيه تكيفي أصغري وفقاً لحجم قائمة الانتظار لكل اتجاه.

تعالج جميع الأساليب السابقة المذكورة أعلاه قضايا محددة مثل التسامح مع الخطأ والإدارة الحرارية. ومع ذلك، فيما يتعلق بالترابط الرأسي المحدود، فإنهم يعتمدون على تعديل الهيكل مثل التسلسل serialization، أو تجميع TSV، أو هيكل الناقل الهجين. من أجل حل هذه المشكلة على بنية عامة وبطريقة أكثر فاعلية، نبدأ من وجهة نظر سياسة التوجيه ونقترح خوارزمية توجيه جديدة لتوزيع حركة المرور.

### 3. الخوارزمية المقترحة (خوارزمية التوجيه التكيفي الموزونة لتوزيع حركة المرور)

#### أ. مخطط توزيع حركة المرور

من أجل توزيع حمل حركة المرور بشكل أكثر فاعلية على الروابط الرأسية المحدودة، نحتاج إلى الاستفادة من معلومات حالة حركة المرور. يمكننا استخدام مقاييس مختلفة كمقياس لحالة حركة المرور للشبكة. ومن الأمثلة على ذلك مساحة المخزن المؤقت buffer المتبقية، أو القنوات الافتراضية المتاحة، أو الطلب على شبكة التصالب crossbar، أو مجموعة من هذه العوامل. تتمثل إحدى الطرق الأكثر مباشرة في استخدام مساحة المخزن المؤقت المتبقية المتوفرة في منفذ الإدخال المقابل للقفزة التالية.

كما تمت مناقشته أعلاه، فإن عرض النطاق الترددي للروابط الرأسية في الشبكة ثلاثية الأبعاد منخفض نسبياً مقارنةً بالروابط الأفقية، ولكن لا تأخذ خوارزميات توجيه الأبعاد التقليدية، مثل XYZ، ولا الخوارزميات الأخرى الحالية التي تستهدف توجيه 3D NoC الازدحام المحتمل في الروابط الرأسية بعين الاعتبار. بناءً على هذه الملاحظة، قررنا استخدام مخطط توجيه تكيفي غير اصغري المسار لتوزيع حمل حركة المرور عبر الشبكة خاصة في الاتجاه الرأسي. ومع ذلك، نظرًا لأن التوجيه غير ذي المسار الأدنى



يسمح بالتوجيه الخاطئ (توجيه الحزمة في اتجاه ليس على الحد الأدنى من المسار إلى وجهتها)، فقد يؤدي ذلك إلى مسارات توجيه طويلة غير ضرورية، وبالتالي زيادة زمن الانتقال وكذلك تقليل الإنتاجية. وبالتالي، تصيح المشكلة التي يتعين حلها تقليل زمن الوصول وتعظيم الإنتاجية مع الاعتراف بسوء التوجيه.

نظرًا لأننا لا نعلم حمل حركة مرور الشبكة العام والتأثير الدقيق لتوزيع حمل حركة المرور على زمن انتقال الشبكة والإنتاجية، فإننا نحاول ببساطة توزيع حمل حركة المرور محليًا مع مراعاة الاتجاه نحو الوجهة. هذا لتقليل الازدحام خاصة في الاتجاه الرأسي وفي نفس الوقت لتجنب مسارات التوجيه الطويلة غير الضرورية بسبب سوء التوجيه. لتنفيذ ذلك، في كل جهاز توجيه، نقوم بتعيين وزن مختلف لكل اتجاه لحساب الأولوية بناءً على حركة المرور. ثم يتم اتخاذ قرار التوجيه وفقًا للأولوية. بالنسبة لتخصيص الوزن في الخوارزمية الخاصة بنا، هناك ثلاث حالات يجب مراعاتها: (1) الاتجاه الرأسي (لا نسمح بالتوجيه الخاطئ في الاتجاه الرأسي)، (2) الاتجاه الأفقي الذي يكون على مسار أدنى (3) الاتجاه الأفقي الذي لا يقع على مسار الحد الأدنى (أي التضليل).

### الخوارزمية 1: تحديد الأوزان

المدخلات:  $X_{di}$ ,  $Y_{di}$ ,  $Z_{di}$

المخرجات: الوزن لكل اتجاه

1:  $X_{di} = X_c - X_{des}$ ; // حساب البعد للوصول للهدف من العقدة الحالية في الاتجاه الأفقي

2:  $Y_{di} = Y_c - Y_{des}$ ;

3:  $Z_{di} = Z_c - Z_{des}$ ; // حساب البعد للوصول للهدف من العقدة الحالية في الاتجاه العمودي

4: initiate all Weights to 0; // ضبط جميع الأوزان للصفر

5: **if** ( $-1 \leq X_{di} \leq 1$  **and**  $-1 \leq Y_{di} \leq 1$  **and**  $-1 \leq Z_{di} \leq 1$ ) **then**

6:     **if** ( $Z_{di} > 0$ ) **then**

7:         Down\_weight = vertical\_close

8:     **else if** ( $Z_{di} < 0$ ) **then**

9:         Up\_weight = vertical\_close;

10:    **end if**

11:    **if** ( $Y_{di} > 0$ ) **then**

12:         South\_weight = horizontal\_close;

```

13: else if (Ydi < 0) then
14:     North_weight = horizontal_close;
15: end if
16: if (Xdi > 0) then
17:     East_weight = horizontal_close;
18: else if (Xdi < 0) then
19:     West_weight = horizontal_close;
20: end if
21: else
22:   if (Zdi > 0) then
23:     Down_weight = vertical_far;
24:   else if (Zdi < 0) then
25:     Up_weight = vertical_far;
26:   end if
27:   if (Ydi > 0) then
28:     South_weight = horizontal_far_min;
29:     North_weight = horizontal_far_detour;
30:   else if (Ydi < 0) then
31:     North_weight = horizontal_far_min;
32:     South_weight = horizontal_far_detour;
33:   end if
34:   if (Xdi > 0) then
35:     East_weight = horizontal_far_min;
36:     West_weight = horizontal_far_detour;
37:   else if (Xdi < 0) then
38:     West_weight = horizontal_far_min;
39:     East_weight = horizontal_far_detour;
40:   end if
41: end if

```

أثناء التنفيذ والتجارب، لاحظنا أن تطبيق نفس استراتيجية التوجيه ذات المسار غير الاصغري على طول الطريق إلى الوجهة لم يعطي نتيجة أداء جيدة، لا سيما في وجود بعض النقاط الفعالة في الشبكة. وهو معروف وشائع في العديد من أنظمة العالم الحقيقي متعددة النواة real-world many-core systems [30]. على وجه الخصوص، عندما تكون نقطة فعالة بالقرب من الوجهة، سيكون من المحتمل جداً أن تختار الحزم

مسارًا غير بسيط وتتجول حول العقدة الوجهة، مما سينشر الازدحام على نطاق قريب من الوجهة ويمنع الحزم الأخرى التي تريد المرور من خلالها. لذلك، فإننا نأخذ المسافة من العقدة الحالية إلى الوجهة في الاعتبار. عندما يتم توجيه الحزمة بالفعل بالقرب من وجهتها، فإننا نمنع توجيهها بشكل خاطئ عن طريق ضبط تخصيص الوزن.

في الخوارزمية 1، نشير إلى فرق الاحداثيات بين العقدة الحالية والوجهة بـ  $X_{di}$  و  $Y_{di}$  و  $Z_{di}$ . يتم تعيين الأوزان الخاصة بـ  $horizontal\_close$  و  $vertical\_close$  على التوالي للاتجاه الرأسي والاتجاه الأفقي عندما تكون العقدة الحالية قريبة من الوجهة. في عملنا، يتم التعامل مع الحالة التي تكون فيها الوجهة ضمن قفزة واحدة في كل بُعد (أي ثلاث قفزات على الأكثر في المجموع إلى الوجهة) من العقدة الحالية على أنها قريبة "close". يتم تعيين الأوزان  $vertical\_far\_min$  و  $horizontal\_far\_min$  على التوالي للاتجاه الرأسي والاتجاه الأفقي على المسار الأصغر عندما تكون العقدة الحالية بعيدة عن الوجهة. يُخصص وزن  $horizontal\_far\_detour$  للاتجاه الأفقي على مسار غير الأصغر من أجل تضليل الحزمة. ولكن بغض النظر عن "far" أو "close"، فإننا لا نسمح بالتوجيه الخاطئ في الاتجاه الرأسي، حيث سيؤدي ذلك إلى زيادة كبيرة بسبب النطاق الترددي المحدود للإرسال الرأسي.

أثناء التجارب، وجدنا أن أداء الشبكة يعتمد بشكل كبير على قيم الوزن للحالة "بعيدًا عن الوجهة". بناءً على مقارنة الإنتاجية بين مجموعات الأوزان المختلفة، فقد حصلنا على مجموعة من قيم الوزن التي تعطي أفضل أداء. سيتم عرض العلاقة بين توزيع الوزن والأداء وتحليلها في القسم الرابع.

### **ب. تجنب الجمود**

بشكل عام، يتطلب التوجيه التكيفي تصميمًا دقيقًا لخوارزمية التوجيه لتجنب الجمود. يحدث الجمود في شبكة الاتصال البيئي عندما يكون هناك تبعية دورية لموارد مثل المخازن المؤقتة والقنوات. في عملنا، نستخدم القنوات الافتراضية بالإضافة إلى مخطط رقم انعكاس البعد (DR) number scheme dimension reversal [31] لتجنب

التبعية الدورية. وبالتالي سيتم تخصيص رقم ( $DR \#$ ) لكل حزمة، وهو عدد المرات التي تم فيها توجيه الحزمة من بُعد أعلى إلى بُعد أقل. يتم تحديد  $DR\#s$  على النحو التالي:

- (1) تمت تهيئة  $DR\#s$  لجميع الحزم إلى 0
- (2) في كل مرة يتم توجيه حزمة من قناة في البعد  $D_i$  إلى قناة في البعد  $D_j$ ، إذا كانت  $j > i$ ، يتم زيادة  $DR\#$ .

استنادًا إلى حقيقة أننا نعتبر الاتجاه الرأسي أولاً، يكون ترتيب الأبعاد هو  $Z$  و  $Y$  و  $X$ . وتنقسم جميع القنوات الافتراضية أيضًا إلى فئات من 0 إلى  $r$  (توجد  $r+1$  قناة افتراضية في كل منفذ)، حيث  $r$  الحد الأعلى لعدد انعكاسات الأبعاد المسموح بها. قد يتم توجيه الحزم التي تحتوي على  $DR\# < r$  إلى أي اتجاه ولكن يجب أن تستخدم قناة افتراضية من الفئة  $DR\#$ . بمجرد أن تقوم الحزمة باستخدام كل انعكاسات أبعادها، مما يعني  $DR\# = r$ ، يجب أن تبدأ في القيام بتوجيه ترتيب البعد من خلال القناة الافتراضية  $r$ .

### ج. حساب حالة حركة المرور

في مخططنا، يتم قياس حركة المرور في كل اتجاه من خلال مساحة المخزن المؤقت المتبقية في العقدة المجاورة. يتم إرسال المعلومات في كل دورة من خلال اتصال إضافي بين العقد المجاورة. نظرًا لأن كل قناة افتراضية في نظامنا تحتوي على سعة تخزين لأربع فتحات، فإن عرض هذا الاتصال الإضافي هو 2 بت. بعد الحصول على معلومات مساحة المخزن المؤقت المجانية، نقوم بحساب حالة حركة المرور للاتجاه المقابل. هنا، نستخدم دالة مباشرة معطاة بواسطة

$$\text{حالة حركة المرور} = \text{مساحة التخزين المؤقت الخالية} \times \text{الوزن المخصص (1)}$$

### د. وصف الخوارزمية العام

يتم تحديد اتجاه توجيه الحزمة في العقدة بناءً على قيمة حالة حركة المرور (تكون الأولوية القصوى للاتجاه الذي يحتوي على أكبر قيمة لشرط حركة المرور) كما هو موضح في الخوارزمية 2.

**الخوارزمية 2:** خوارزمية التوجيه التكيفية لتوزيع حركة المرور في شبكات ثلاثية

الابعاد

Input: Traffic, (Xdi, Ydi, Zdi), DR# //مدخلات الخوارزمية

Output: Next Hop (E, W, N, S, U, D, L) // الخرج هو الفقرة التالية

1: **if** (Xdi = Ydi = Zdi = 0) **then**

2: Deliver the packet to the local node and **exit**; // عند الوصول للهدف

3: **end if**

4: **if** (DR#<r) **then**

5: *next\_hop* = Best (Traffic condition of all directions);

6: **else if** (DR#=r) **then**

7: **return** lowest dimension differs from the destination;

8: **end if**

في الخوارزمية، يعتمد الاتجاه الذي يجب أن تجتازه حزمة البيانات على فرق التنسيق بين العقدة الحالية والعقدة الوجهة، DR# للحزمة، بالإضافة إلى حالة المرور المحسوبة لجميع الاتجاهات الممكنة. في الحالات العامة، تأخذ الخوارزمية في الاعتبار أربعة اتجاهات مرشحة باستثناء الاتجاه الذي وصلت منه الحزمة. ومع ذلك، في حالة أن العقدة الحالية هي مصدر الحزمة، نحتاج إلى الاختيار من بين خمسة اتجاهات مرشحة وفقاً لظروف حركة المرور. التابع Best هو اختيار الاتجاه ذي القيمة الأكبر، مما يعني الاتجاه الأقل ازدحاماً، ثم إرجاع هذا الاتجاه (انظر الخوارزمية 3). في الحالات التي يتم فيها تساوي اتجاهين أو أكثر، نختار الاتجاه الأكبر وزناً لمنع الكثير من سوء التوجيه.

### الخوارزمية 3: تابع اختيار الاتجاه الأفضل

*Best ()*

حالات المرور لجميع الاتجاهات المرشحة: الدخل

الاتجاه ذو أفضل حالة حركة مرور: الخرج

1: **for** all candidate directions

2: choose *the direction* with maximum traffic condition value

3: **if** (*the direction* with best traffic is unique) **then**

4: **return** *the direction*;

5: **else if** (two or more directions are tied to be the best) **then**

6: **return** *the direction with biggest weight*;

6: **end if**

7: **end for**

### 4. النتائج التجريبية

#### أ. الإعداد التجريبي

لإثبات فعالية خوارزمية التوجيه التكيفية المقترحة الموضحة أعلاه، قمنا أيضًا بتنفيذ خوارزمية توجيه ZYX التقليدية بالإضافة إلى خوارزمية التوجيه AdaptiveXYZ [21] تم إجراء المحاكاة على محاكي Noxim دقيق الدورة ونمذجة جميع المكونات الرئيسية لهيكل الشبكة ثلاثية الأبعاد وأجرينا محاكاة لقياس زمن الانتقال والإنتاجية للشبكة. كما استخدمنا نفس عدد القنوات الافتراضية لكل خوارزمية على شبكة ثلاثية الأبعاد بأكملها بحجم  $4 \times 4 \times 4$ . بالنسبة لجميع أجهزة التوجيه، قمنا بتعيين عرض القناة الأفقية على 32 بت وعرض القناة الرأسية على 8 بت ( $4/1$  عرض القناة الأفقية). في تطبيقنا، يحتوي كل منفذ على أربع قنوات افتراضية، مما يعني أن كل حزمة يمكن أن تحتوي على توجيه انعكاس البعد ثلاث مرات على الأكثر. سعة المخزن المؤقت FIFO لكل قناة افتراضية هي أربع وحدات.

كمقياس للأداء، نستخدم زمن الانتقال المحدد على أنه عدد الدورات بين بدء إرسال رسالة صادرة عن عنصر معالجة في العقدة المصدر والوقت الذي يتم فيه تسليم الرسالة أخيراً إلى العقدة الوجهة. كمقياس أداء آخر، نستخدم معدل نقل الشبكة، والذي يتم قياسه بعدد flits المستلمة على مدار دورات المحاكاة الإجمالية. يتم تسخين جهاز المحاكاة لـ 10000 دورة ثم يتم قياس متوسط الأداء على مدى 100000 دورة لاحقة.

من أجل المحاكاة، استخدمنا ثلاثة أمثلة تركيبية بأنماط حركة مرور مختلفة: حركة مرور عشوائية موحدة uniform random traffic، وحركة مرور نقطة فعالة hotspot traffic، وحركة مرور متمم البت bit-complementary traffic. في نمط المرور العشوائي الموحد، تولد كل عقدة رزمًا لكل وجهة باحتمالية متساوية. في نمط مرور النقاط الفعالة، يتم اختيار عقدة أو أكثر كنقاط فعالة تتلقى جزءًا إضافيًا من حركة المرور بالإضافة إلى حركة المرور الموحدة. في المحاكاة الخاصة بنا، يتم اختيار العقدة (2، 2، 2) كنقطة ساخنة ويتم توجيهه 15٪ من الرسائل التي تم إنشاؤها إلى نقطة الاتصال. في نمط الحركة لمتتم البت، ترسل كل عقدة حزمًا إلى العقدة التكميلية للبتات الخاصة بها، والتي يتم تحديدها بواسطة مكمل البتات لمعرف العقدة الخاص بها المرقم بترتيب البعد. على سبيل المثال، العقدة (0، 2، 1) ترسل حزمًا إلى العقدة (3، 1، 2). وبالتالي، يجب أن تمر كل حزمة مرسله من خلال اتصال بين الطبقات، والذي يوضح بشكل أفضل تأثير توزيع الحمل في الاتجاه الرأسي.

### ب. تحليل الأداء

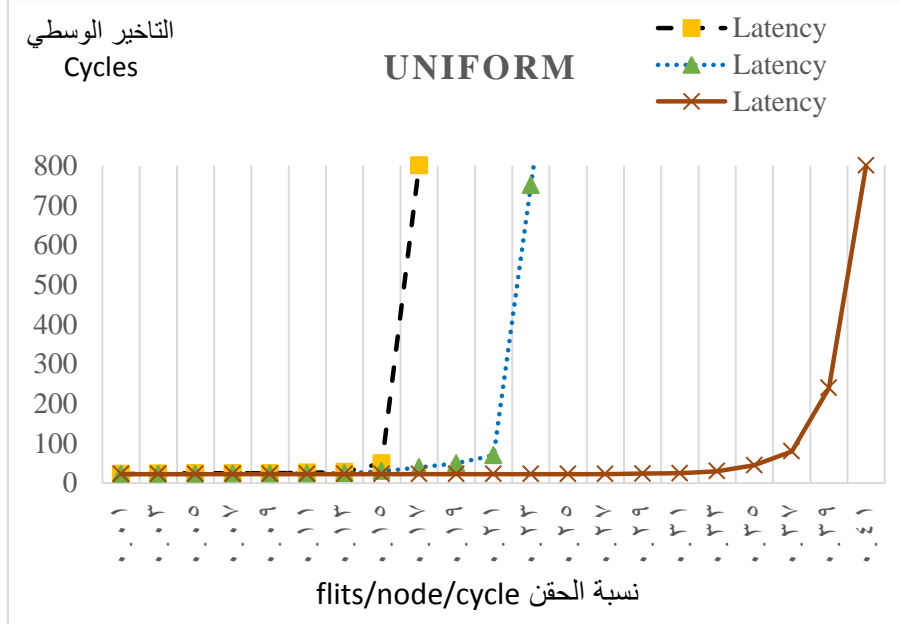
كما هو مبين في الأشكال (7، 6، 5، 4، 9، 8)، في جميع أنماط المرور الثلاثة، تتفوق خوارزمية التوجيه المقترحة على توجيهه ZYX وتوجيهه AdaptiveXYZ في كل من زمن الانتقال والإنتاجية. هذا يأتي من نقطتين. أولاً، من خلال تعيينات الوزن المناسبة، يمكن للخوارزمية المقترحة استخدام الروابط الرأسية بشكل أفضل وتوزيع حركة المرور بشكل أكثر فاعلية عبر الشبكة بأكملها مقارنة بخوارزميات التوجيه الأخرى. علاوة على ذلك، تسمح الخوارزمية الخاصة بنا بتوجيه المسار غير الأدنى، والذي يمكنه توزيع

حركة المرور بشكل أفضل وتجاوز المنطقة المزدحمة في الشبكة بشكل أفضل. يوضح الجدول (1) ملخصاً للنتائج من حيث الحد الأقصى للإنتاجية التي تم الحصول عليها عن طريق زيادة معدل الحقن إلى أقصى حد حتى تشبع الشبكة. من الجدول، يمكننا أن نرى أن الخوارزمية المقترحة لها ميزة كبيرة على النهجين الآخرين. وخصوصاً في حركة المرور التكميلية للبت، نحصل على معدل نقل أعلى بنسبة 105.4% من AdaptiveXYZ نظراً لاستخدام الارتباطات الرأسية بشكل أفضل.

الجدول (1) تحسين الاداء الذي تحققه الخوارزمية المقترحة بالمقارنة مع خوارزمية XYZ

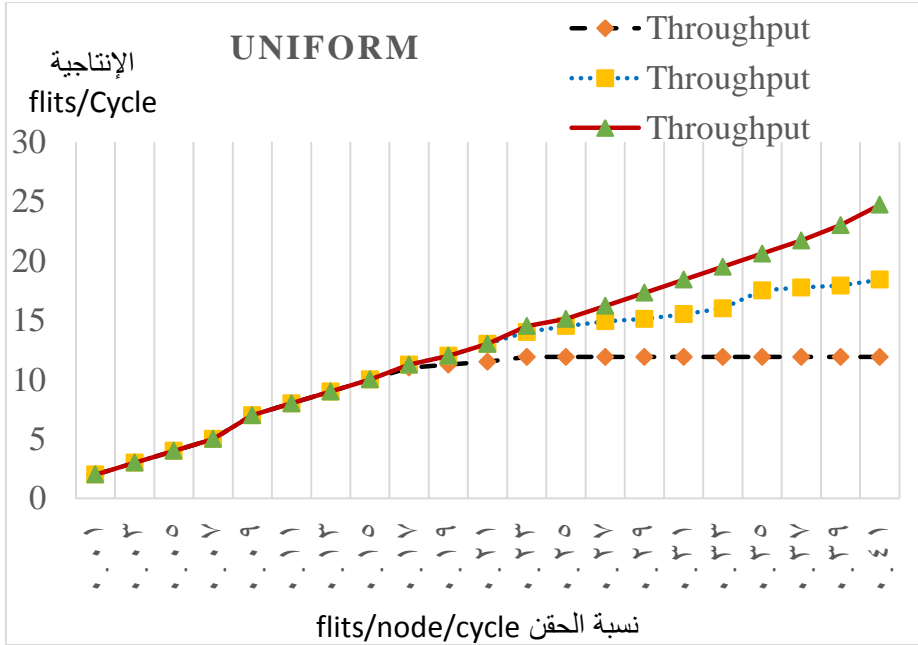
### و Adaptive XYZ

نمط حركة المرور	الإنتاجية flits/Cycle			التحسين بالمقارنة مع	
	ZYX	Adaptive XYZ	proposed	XYZ	Adaptive XYZ
Uniform	11.9	18.4	24.7	107.56%	34.24%
Hotspot	5.4	5.7	6.2	14.81%	8.77%
Bit-Complementary	6.3	7.4	15.2	141.27%	105.4%

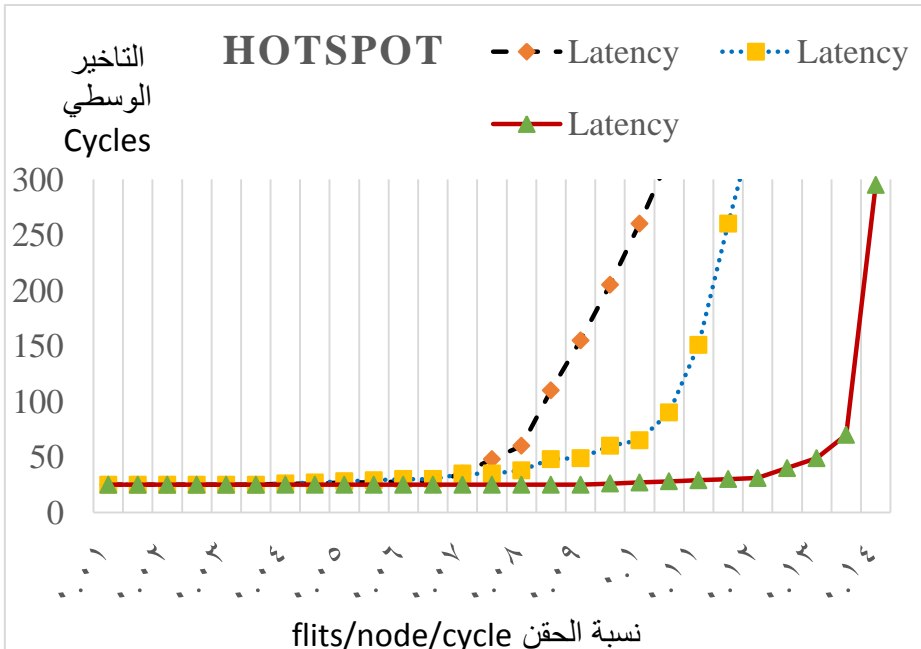


الشكل (4) التأخير الوسطي بالنسبة لملف حركة المرور العشوائية الموحدة

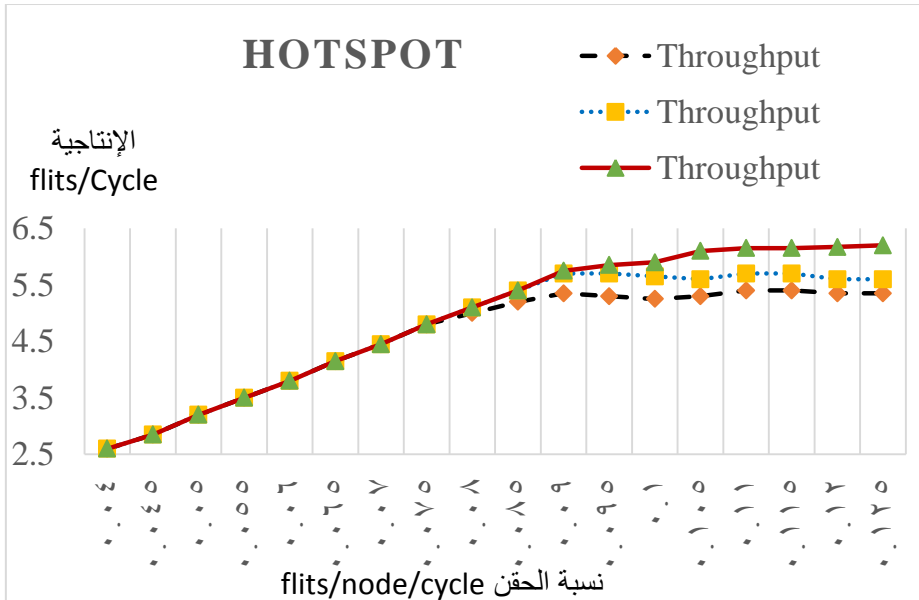




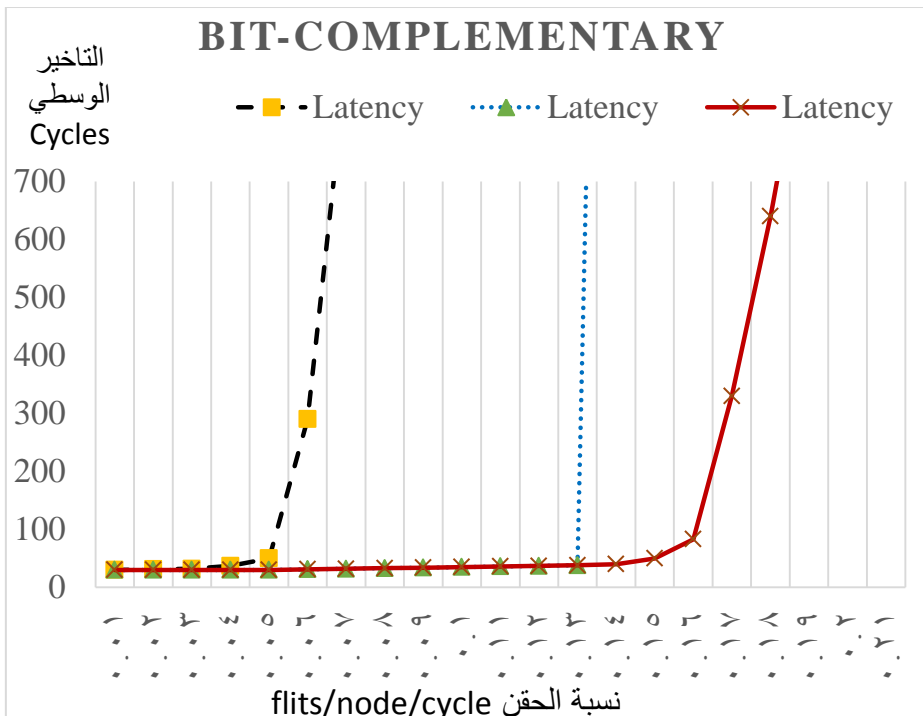
الشكل (5) الإنتاجية بالنسبة لملف حركة المرور العشوائية الموحدة



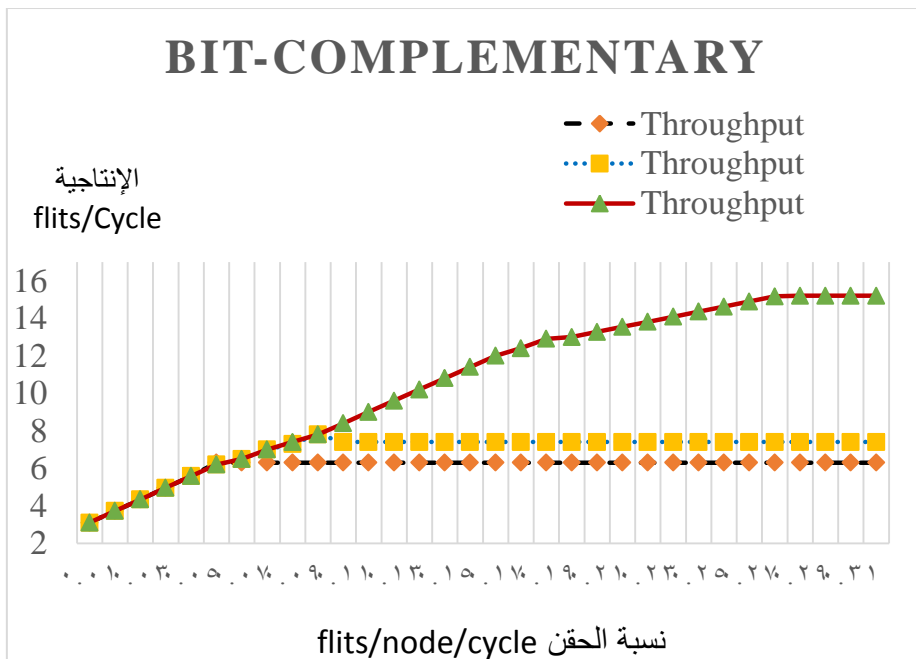
الشكل (6) التأخير الوسطي بالنسبة لملف النقطة الساخنة



الشكل (7) الإنتاجية بالنسبة لملف النقطة الساخنة



الشكل (8) التأخير الواسطي بالنسبة لملف متمم البت



الشكل (9) الإنتاجية بالنسبة لملف مكمل البت

### ج. الحساسية لتخصيص الوزن والتطبيق

كما نوقش أعلاه، يعتمد الأداء على كيفية تعيين الأوزان. ومع ذلك، إذا كان الأداء شديد الحساسية لتباين قيم الأوزان أو لأنماط حركة المرور المختلفة، فسيكون من الصعب تعيين أوزان مناسبة. لقياس الحساسية، قمنا بدراسة الإنتاجية من أجل عدة قيم  $horizontal\_far\_min/horizontal\_far\_detour$  والنسبة  $vertical\_far/horizontal\_far\_detour$ .

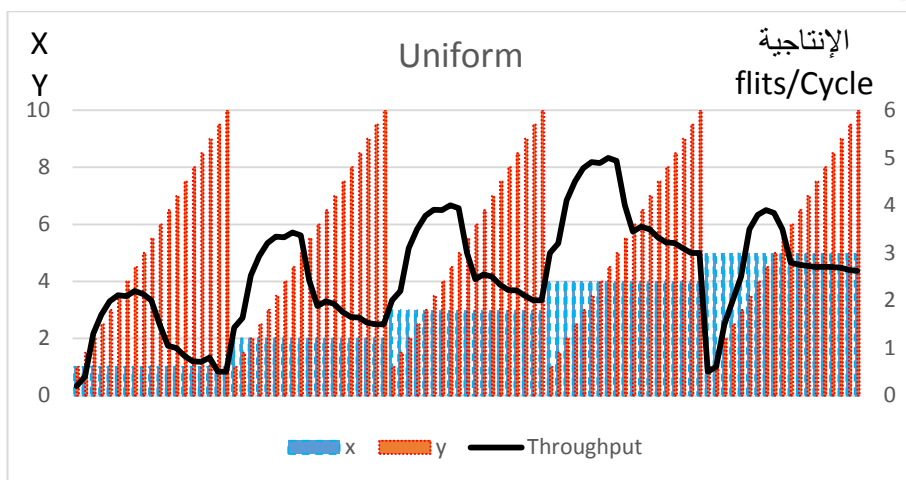
وتمكنا أن نرى أن الإنتاجية منخفضة جداً مع قيم  $x$  الصغيرة، حيث يكون وزن الاتجاهات غير الدنيا (التنافية) كبيراً نسبياً، مما يؤدي إلى الكثير من التضليل في التوجيه. في حالة حركة المرور في النقاط الساخنة وأنماط المرور العشوائية الموحدة، تكون النقطة المثلى عند  $x = 4$  كما هو مبين في الشكل (10) الذي يوضح تغير الإنتاجية بالنسبة لنمط حركة المرور العشوائية الموحد مع تغير كل من النسب:

$$X = \text{horizontal\_far\_min} / \text{horizontal\_far\_detour}$$

$$Y = \text{coordinate is for vertical\_far} / \text{horizontal\_far\_detour}$$

وفي حالة نمط المرور التكميلي للبت، تكون النقطة المثلى عند  $x = 4.5$ ، لكن الأداء عند  $x = 4$  ليس أسوأ بكثير من ذلك عند النقطة المثلى. لذلك اخترنا  $x = 4$  كنسبة وزن لشبكتنا، والتي تساوي في الواقع حجم المخزن المؤقت لكل قناة افتراضية. هذا يعني أنه فقط عندما يتم استخدام المخزن المؤقت في الاتجاه الأدنى، سيتم توجيه الحزمة إلى اتجاه غير محدود، مما يعطي أصغر احتمال لسوء التوجيه. في ظل جميع أنماط الحركة الثلاثة، تكون النقطة المثلى عند  $y = 5.5$ ، مما يعني أننا نعطي أهمية أكبر للروابط الرأسية مقارنة بالروابط الأفقية من أجل توجيه الحزم إلى الاتجاه الرأسي في وقت سابق لتجنب حركة المرور غير المتوازنة على الروابط الرأسية.

الأداء ليس حساساً جداً لتباين الأوزان بالقرب من النقطة المثلى. علاوة على ذلك، فإن النقطة المثلى لنمط مرور واحد توفر أيضاً أداءً جيداً لأنماط حركة المرور الأخرى، مما يعني أن الأداء ليس حساساً جداً لأنماط حركة المرور المختلفة.



الشكل (10) تأثير الإنتاجية بتغيير الأوزان في نمط حركة المرور الموحد

## 5. الاستنتاج

في هذا البحث، اقترحنا توزيع لحركة المرور بتوجيه تكيفي لشبكة ثلاثية الأبعاد ذات عرض نطاق عمودي محدود. من خلال الاستخدام الفعال للروابط الرأسية وتوزيع الأحمال على الشبكة بالكامل، يمكننا تحسين زمن الوصول وإنتاجية الشبكة بشكل كبير. مقارنةً بخوارزمية التوجيه AdaptiveXYZ المقترحة مؤخرًا، حصل المخطط المقترح على ما يقرب من 50% أعلى من إجمالي الانتاجية تحت ثلاثة أنماط حركة، خاصة بالنسبة لنمط المرور التكميلي للبت، حصلت الخوارزمية المقترحة على 105.4% أعلى من إجمالي الانتاجية. تعتبر الخوارزمية المقترحة قوية من حيث أن الأداء غير حساس لأنماط حركة المرور المختلفة. استخدمت محاكاتها فقط بنية ربط شبكية ثلاثية الأبعاد عادية، والتي تعتمد على توصيل العقدة بالعقدة. ومع ذلك، فهي لا تقتصر على الشبكة فقط ويمكن تطبيقها على طوبولوجيا أخرى مثل الهيكل الهجين للناقل عن طريق تعديل حساب الحركة العمودية لحساب قيمة ضغط الناقل [29] [21].

## REFERENCES:

- [1] Li, C.-L.; Yoo, J.-C.; Han, T.H. Energy-efficient custom topology-based dynamic voltage-frequency island-enabled network-on-chip design. *J. Semicond. Technol. Sci.* **2018**, *18*, 352–359.
- [2] Fang, J.; Zong, H.; Zhao, H.; Cai, H. Intelligent mapping method for power consumption and delay optimization based on heterogeneous NoC platform. *Electronics* **2019**, *8*, 912. 3. Lee, Y.S.; Kim, S.; Han, T.H. Aging-resilient topology synthesis of heterogeneous manycore network-onchip using genetic algorithm with flexible number of routers. *Electronics* **2019**, *8*, 1458.
- [3] Lee, Y.S.; Kim, S.; Han, T.H. Aging-resilient topology synthesis of heterogeneous manycore network-onchip using genetic algorithm with flexible number of routers. *Electronics* **2019**, *8*, 1458.
- [4] Taheri, E.; Mohammadi, K.; Patooghy, A. On-off: A reactive routing algorithm for dynamic thermal management in 3d NoCs. *IET Comput. Digit. Tech.* **2018**, *13*, 11–19.
- [5] Chao, C.-H.; Chen, K.-C.; Yin, T.-C.; Lin, S.-Y.; Wu, A.-Y. Transport-layer-assisted routing for runtime thermal management of 3D NoC systems. *ACM Trans. Embed. Comput. Syst.* **2013**, *13*, 1–22.
- [6] Chen, K.-C.; Kuo, C.-C.; Hung, H.-S.; Wu, A.-Y.A. Traffic-and thermal-aware adaptive beltway routing for three dimensional network-on-chip systems. In *Proceedings of the 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, Beijing, China, 19–23 May **2013**; pp. 1660–1663.
- [7] WeldeZion, A.Y.; Ebrahimi, M.; Daneshtalab, M.; Tenhunen, H. Automated power and latency management in heterogeneous 3d nocs. In *Proceedings of the 8th International Workshop on Network on Chip Architectures*, Waikiki, HI, USA, 5 December **2015**; pp. 33–38.
- [8] Jiang, X.; Lei, X.; Zeng, L.; Watanabe, T. Fully adaptive thermal-aware routing for runtime thermal management of 3D network-on-chip. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Hong Kong, China, 16–18 March **2016**.

- [9] Lee, Y.-S.; Hsin, H.-K.; Chen, K.-C.; Chang, E.-J.; Wu, A.-Y.A. Thermal-aware dynamic buffer allocation for proactive routing algorithm on 3D network-on-chip systems. In Proceedings of the Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test, Hsinchu, Taiwan, 28–30 April **2014**; pp. 1–4.
- [10] Salamat, R.; Khayambashi, M.; Ebrahimi, M.; Bagherzadeh, N. LEAD: An adaptive 3D-NoC routing algorithm with queuing-theory based analytical verification. *IEEE Trans. Comput.* **2018**, *67*, 1153–1166.
- [11] Seiculescu, C.; Murali, S.; Benini, L.; De Micheli, G. SunFloor 3D: A tool for networks on chip topology synthesis for 3-D systems on chips. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **2010**, *29*, 1987–2000.
- [12] Charif, A.; Coelho, A.; Ebrahimi, M.; Bagherzadeh, N.; Zergainoh, N.-E. First-last: A cost-effective adaptive routing solution for tsv-based three-dimensional networks-on-chip. *IEEE Trans. Comput.* **2018**, *67*, 1430–1444.
- [13] Zou, T.; Zhang, C.; Peng, X.; Peng, Y. An inter-layer-distance based routing algorithm for 3D network-onchip. In Proceedings of the CCF National Conference on Computer Engineering and Technology, Yinchuan, China, 15–17 August **2018**; pp. 26–37.
- [14] V.F. Pavlidis and E.G. Friedman, *Three-dimensional Integration Circuit Design*, Morgan Kaufmann, **2008**.
- [15] A.-M Rahmani et al., “Power and area optimization of 3D networks-onchip using smart and efficient vertical channels,” in Proc. PATMOS, **2011**, pp. 278-287.
- [16] H. Sangki, “3D super-via for memory applications,” *Micro-Systems Packaging Initiative (MSPI) Packaging Workshop*, **2007**.
- [17] C. Liu et al., “Vertical interconnects squeezing in symmetric 3D mesh Network-on-chip,” in Proc. ASP-DAC, **2011**, pp. 357-362.
- [18] K. Puttaswamy and G.H. Loh, “Thermal herding: microarchitecture techniques for controlling hotspots in high-performance 3D-integrated processors,” in Proc. ISCA, **2008**, pp. 251-261.

- [19] I. Loi, S. Mitra, T.H. Lee, S. Fujita, and L. Benini, "A low-overhead fault tolerance scheme for TSV-based 3D network on chip links," in Proc. ICCAD, **2008**, pp. 598-602.
- [20] S. Pasricha, "Exploring serial vertical interconnects for 3D ICs," in Proc. DAC, 2009, pp.581-586.
- [21] A.-M. Rahmani et al., "ARB-NET: a novel adaptive monitoring platform for stacked mesh 3D NoC architectures," in Proc. ASP-DAC, **2012**, pp. 413-418.
- [22] Intel Corporation, "A touchstone delta system description," in: Intel Advanced Information, **1991**.
- [23] C.J. Glass and L.M. Ni, "The turn model for adaptive routing," in Proc. ISCA, **1992**, pp. 278-287.
- [24] M. Li, Q. Zeng, and W. Jone, "DyXY – a proximity congestion-aware deadlock-free dynamic routing method for network on chip," in Proc. DAC, **2006**, pp. 849-852.
- [25] G. Ascia, V. Catania, M. Palesi, and D. Patti, "Implementation and analysis of a new selection strategy for adaptive routing in networks-on-chip", IEEE Trans. Comput., vol. 57, no. 6, pp. 809-820, June. **2008**.
- [26] S. Ma, N.E. Jerger, and Z. Wang, "DBAR: an efficient routing algorithm to support multiple concurrent applications in networks-on-chip", in Proc. ISCA, **2011**, pp. 413-424.
- [27] M. Ebrahimi et al., "CATRA- congestion aware trapezoid-based routing algorithm for on-chip networks," in Proc. DATE, **2012**, pp. 320-325.
- [28] S.Y. Lin et al., "Traffic-and thermal aware routing for throttled threedimensional network-on-chip systems," in Proc. VLSI-DAT, **2011**, pp. 1-4.
- [29] A.-M. Rahmani et al., "Congestion aware, fault tolerant, and thermally efficient inter-layer communication scheme for hybrid NoC-bus 3D architectures," in Proc. NOCS, **2011**, pp. 65-72.
- [30] G.J. Pfister and V.A. Norton, "Hotspot contention and combining in multistage interconnection networks," IEEE Trans.Comput., vol. 34, no. 10 ,pp. 943-948, 1985.



- [31] W.J. Dally and C.L. Seitz, "Deadlock-free adaptive routing in multicomputer networks using virtual channels," IEEE Trans.Parallel Distrib. Syst., vol. 1, no. 3, pp. 187-196, Oct.

