
6 Resümee und Ausblick

In der Einleitung (Kapitel 1) wurde bereits ein ausführlicher Überblick über den Inhalt der Kapitel 2 (Problemstellung), 3 (Stand der Technik) und 4 (Timing-driven Floorplanning im PLAYOUT-System) gegeben, so daß sich hier darauf beschränkt werden kann, die Resultate der Messungen und Experimente (Kapitel 5) zusammenzufassen. Verbunden damit ist eine kritische Bewertung und ein Ausblick auf interessante Fragestellungen und offene Forschungsthemen im weiteren Kontext dieser Arbeit.

Die Messungen wurden hauptsächlich an drei Chipentwürfen durchgeführt, die aus dem Volumenvisualisierungsprojekt ([Lic98]) der AG Zimmermann stammen und deren Entwurf zuvor mit kommerziellen Tools geschah. Sie weisen einige gemeinsame Charakteristika auf:

- die Verwendung der ES2_0.7-Technologie (Sub-Micron-Bereich)
- eine geringe bis mittlere Komplexität gemessen an der Anzahl von Subzellen
- eine hohe Komplexität gemessen an der Anzahl von Netzen
- eine moderate Chipgröße im Bereich von 6 mm · 6 mm bis 8 mm · 8 mm
- die maximale Verzögerungszeit des Chips ist weitgehend bestimmt durch die abgeschätzten Verzögerungszeiten der Subzellen.

Um darüber hinaus zu Aussagen über den Deep Sub-Micron-Bereich zu gelangen, wurden bei den Messungen mit verschiedenen Skalierungen experimentiert und eventuell zusätzliche Modifikationen an den Eingabebeispielen vorgenommen:

- Skalierung der ES2_0.7-Technologie auf technologische Parameter (Widerstände, Kapazitäten, Verzögerungszeiten der Subzellen), die eine Technologie mit einem Drittel der Strukturweiten haben könnte (SK2, SK3)
- Berücksichtigung der stark unterschiedlichen Ansetzung des Wertes für die spezifische Leitungskapazität in der Literatur (SK1, SK3)
- Begrenzung der abgeschätzten Verzögerungszeiten der Subzellen, um den Einfluß der Verzögerungszeiten der CuD-Netze auf die Gesamtverzögerungszeit zu erhöhen (*grp.2* statt *grp*, *lon.2* statt *lon*)
- Vergrößerung der Dimensionen der Subzellen um den Faktor 5.

Bei der Durchführung der Messungen stand zusätzlich eine Reihe von Freiheitsgraden zur Verfügung, deren verschiedene Einstellmöglichkeiten kombinatorisch explodieren:

- Wahl des Platzierungsalgorithmus (TP, TPmK, TPmB aus dem *planner.v6* oder Einlesen des Ergebnisses der Quadropartitionierung des *planner.v5*)
- Einstellen der Parameter der Platzierungsalgorithmen und der darunter liegenden Mincut-Partitionierungsverfahren (z. B. Ratiocut oder Mincut als Kostenfunktion, Schärfe der Balanzebedingung, Anzahl zufälliger Startwerte, Tiefe des Look-ahead bei der Festlegung der Orientierung einer Slicinglinie, u.v.m.)
- Wahl des Modus der globalen Verdrahtung (pdGR, pdGR-N, pdGR-A)
- Ein- oder Ausschalten der Überzellverdrahtung
- Einstellen der Parameter der globalen Verdrahtung (Gewichtungsfaktoren für Freiflächen, reservierte Flächen und Überlastflächen, prozentuale Vorbelegung bei Überzellverdrahtung).

Außerdem sollte man bei der Bewertung der Ergebnisse nicht vergessen, daß die meisten Verfahren Heuristiken oder gar Kombinationen von Heuristiken sind, so daß einzelne Ausreißer in den Ergebnissen eventuell nur auf einer ungünstigen Konstellation beruhen und wenig Aussagekraft über die allgemeine Anwendbarkeit des Verfahrens haben. So wurde anhand der Abbildungen 5.18 und 5.19 ausführlich ein Beispiel dokumentiert, in dem die ungünstige Anwendung einer Steinerbaumheuristik für ein einziges zeitkritisches Netz das erwartete Resultat beim Vergleich zweier Algorithmen auf den Kopf stellte.

Nach diesen Vorbemerkungen sind folgende Ergebnisse bemerkenswert:

- die eingeschränkte Variabilität der Entwürfe in Bezug auf die maximale Verzögerungszeit im Vergleich zu den Unterschieden bezüglich der Chipfläche
- die große Ausdehnung des Lösungsraums, in dem die verschiedenen Algorithmen gute Lösungen mit einem akzeptablen Trade-Off zwischen maximaler Verzögerungszeit und Chipfläche finden
- die Robustheit der Algorithmen gegenüber geringfügigen Modifikationen der Eingabeparameter
- die größere Korrelation von maximaler Verzögerungszeit und Gesamtnetzlänge gegenüber derjenigen von maximaler Verzögerungszeit und Chipfläche
- die relativ konstante Größe des Einsparpotentials bei OTC-Verdrahtung, unabhängig davon, ob die Entwürfe nur aus flexiblen Subzellen bestehen oder eine Mischung aus flexiblen Subzellen und Makrozellen sind.

Tabelle 6.1: Variabilität der maximalen Verzögerungszeit und der Chipfläche

Technologie	Beispiel	maxDel min	maxDel max	%	area min	area max	%
ES2_0.7	adr ¹	13,83	14,25	3,0	92,0	99,9	8,6
	adr ²	13,74	13,96	1,6	83,4	94,3	13,1
	grp.2	24,86	25,15	1,2	89,0	103,6	16,4
	lon.2	11,53	11,67	1,2	95,4	112,4	17,8
SK1	adr	17,02	18,30	7,5	92,2	99,4	7,8
	grp.2	25,32	26,35	4,1	89,2	113,7	27,5
	lon.2	12,70	13,48	6,1	87,0	111,2	27,8
SK2	adr	4,901	5,315	8,4	91,8	99,0	7,8
	grp.2	8,37	8,50	1,6	88,8	112,0	26,1
	lon.2 ¹	4,041	4,225	4,6	82,9	115,3	39,1
	lon.2 ²	3,979	4,210	5,8	76,1	110,4	45,1
SK3	adr	6,700	8,005	19,5	91,7	99,0	8,0
	grp.2 ¹	8,56	9,53	11,3	87,1	112,8	29,5
	grp.2 ²	8,50	9,04	6,4	77,6	101,3	30,5
	lon.2	4,871	5,248	7,7	83,8	110,4	31,7
ACMOS4h	minproc	14,77	14,97	1,4	83,4	94,1	12,8

Eine zusammenfassende Gegenüberstellung der Ergebnisse bezüglich der Variabilität von maximaler Verzögerungszeit und Chipfläche gibt Tabelle 6.1. Die prozentualen Angaben geben jeweils die Abweichung des schlechtesten vom besten Entwurf an (wobei zielgerichtet schlecht entworfene und Ausreißer nicht berücksichtigt sind), die in den Spalten **min** und **max** davor aufgeführt sind. Für Eingabebeispiele mit hochgestellter Eins (¹) sind die Ergebnisse ohne OTC-Verdrahtung aus Kapitel 5.4 die Grundlage, bei hochgestellter Zwei (²) sind es die Ergebnisse mit OTC-Verdrahtung aus Kapitel 5.4 und ansonsten die Ergebnisse ohne OTC-Verdrahtung aus Kapitel 5.3.

Trotz teilweise bedeutender Differenzen in den Chipflächen liegen die maximalen Verzögerungszeiten meist recht nahe beieinander. Dieses läßt darauf schließen, daß die (guten) Platzierungen die den Schaltungen inhärente Lokalität finden und ausnutzen. Die Verschnittflächen, die durch ungünstige Orientierung der Slicinglinien in der Nähe von Makrozellen hervorgeru-

fen werden, sind verantwortlich dafür, daß die Chipflächen mancher (guter) Entwürfe um mehr als die üblichen 10 % von der minimalen Chipfläche abweichen. Da das Ziel dieser Arbeit aber nicht darin liegt, den jeweils optimalen Entwurf zu finden, sind diese Abweichungen nicht abträglich, sondern verhelfen uns zu inhomogenen Lösungen und damit allgemeineren Aussagen.

Vielmehr stellt sich an dieser Stelle die Frage nach dem Wert der erhaltenen Abschätzungen für den hierarchischen Entwurfsablauf. Dadurch daß das Zeitverhalten auf eine einzige Zahl (oder ein Paar: $\max\Delta t$, $\min\Delta t$) komprimiert wird, bekommt die Planung nicht immer eine aussagekräftige Information. Hat man für jede Subzelle verschiedene Alternativen (gewichtete Timinggraphen) mit unterschiedlich abgeschätzten Verzögerungszeiten, so ergibt die Kombination der Minima nicht unbedingt das beste Zeitverhalten der CuD. Dieses bekommt man nur dann garantiert, wenn Timinganalysen der CuD mit allen Kombinationen der gewichteten Timinggraphen der Subzellen ausgeführt werden - ein wahrhaft aufwendiges Unterfangen. (Deshalb beschränkte sich diese Arbeit im praktischen Teil auch darauf, nur je einen gewichteten Timinggraphen pro Subzelle zu verwenden.)

Der eigentliche Wert liegt in der Unterscheidung der guten von den schlechten Entwürfen. Würde man den hierarchischen zeitgesteuerten Entwurf konsequent mit der konzeptionellen Erweiterung der Shape Functions zu SD-Kurven durchführen, so würde die SD-Kurve nur die guten Entwürfe repräsentieren. Die Sizingsschritte nach Platzierung und globaler Verdrahtung wären mit der Auswahl des entsprechenden gewichteten Timinggraphen zum Kurvenpunkt verbunden. Da die Abschätzungen der Verzögerungszeiten wenig differieren, kann man davon ausgehen, daß auch die Auswahl unterschiedlich gewichteter Timinggraphen zu verschiedenen Kurvenpunkten nur eingeschränkte Auswirkungen hat. Ähnlich wie bei der Platzierung mittels Terminal-Propagierung für die Berechnung der Pseudotopographie ein Stellvertreter für die Formalalternativen jeder Subzelle ausgewählt wird, der nur selten mit dem tatsächlich ausgewählten Kurvenpunkt zusammenfällt, so kann die zeitgesteuerte Platzierung bei der Auswahl eines guten Repräsentanten für die gewichteten Timinggraphen einer Subzelle hoffen, die maximale Verzögerungszeit der CuD hinreichend genau vorhersagen und damit berücksichtigen zu können. Da eine gewisse Korrelation zwischen Chipfläche und maximaler Verzögerungszeit nachgewiesen wurde, gibt es mehrere in Frage kommende Strategien zur Bestimmung eines guten Repräsentanten für den gewichteten Timinggraphen einer Subzelle:

- denjenigen mit der geringsten maximalen Verzögerungszeit
- denjenigen zum Kurvenpunkt mit der geringsten Fläche
- denjenigen zum Kurvenpunkt mit dem besten Gütemaß.

Je größer der Einfluß der CuD-Netze auf die Gesamtverzögerungszeit ist, desto weniger Auswirkungen hat es, wenn der Repräsentant und der ausgewählte gewichtete Timinggraph geringfügige Differenzen aufweisen.

Im theoretischen Teil der Arbeit zur zeitgesteuerten Plazierung wurde herausgestellt, daß es bei der Kombination von Netzgewichtung und Mincut-Partitionierung mit Terminal-Propagierung (TPmK) vier Gewichtungsfaktoren gibt: Einen für zeitkritische interne Netze, einen für unkritische interne Netze, einen für zeitkritische externe Netze und einen für unkritische externe Netze. Diese an sich selbstverständliche Beobachtung findet man in der Literatur selten erwähnt. Der Test mehrerer Parametersätze (nicht näher dokumentiert in dieser Arbeit) legt den Schluß nahe, daß der Algorithmus robust auf die Parameterwahl reagiert, solange die Werte vernünftig gewählt sind, d. h. zeitkritische Netze stärker gewichtet werden als unkritische und externe Verbindungen nicht stärker als interne.

Mit der Budgetierung (TPmB) wurde ein interessanter Ansatz zur zeitgesteuerten Plazierung vorgestellt, der "schwache" Restriktionen verwendet. Die Grundidee besteht darin, daß bei der in Breitensuche durchgeführten Partitionierung möglichst frühzeitig Verletzungen von Zeitvorgaben erkannt werden sollen. Hierzu wird jede Subzelle, an die ein Netz mit propagierten Terminals anschließt, probenhalber in die von diesem Terminal entfernte Partition gesteckt und durchgerechnet, ob die veranschlagte Verzögerungszeit noch einzuhalten ist. Ansonsten wird ein Bindungswunsch an die nahe Partition generiert. Die Konfliktlösung bestimmt, ob aus dem Bindungswunsch eine feste Bindung (der Subzelle an die gewünschte Partition) resultiert oder nicht. Die Auflistung aller oder nur der nicht erfüllten Bindungswünsche gibt dem Entwerfer einen exzellenten Überblick über die Stellen, an denen der Entwurf scheitern könnte oder nachoptimiert werden muß.

Überraschend bei der Durchführung der Versuche war es, daß die beiden Verfahren TPmK und TPmB stets unterschiedliche Lösungen fanden. Eine deutliche Tendenz, die die Bevorzugung des einen oder des anderen Verfahrens begründen würde, gab es dabei nicht. (Die Stichprobengröße wäre auch zu gering, um dazu eine fundierte Aussage zu treffen.)

Abgesehen von der (behebaren) Schwäche mit der Orientierung von Slicinglinien in der Nähe von Makrozellen halten sich TPmB und TPmK auch bei der Minimierung der Chipfläche sehr beachtlich im Vergleich zur Quadropartitionierung. Umgekehrt zeigt die Quadropartitionierung teilweise erstaunlich gute Resultate bei der maximalen Verzögerungszeit, manchmal macht sich allerdings schon bemerkbar, daß sie diese bei der Optimierung nicht berücksichtigt, z. B. stammt die um 19,5 % höhere maximale Verzögerungszeit beim Beispiel *adr* (SK3) von einer Quadropartitionierung.

Durch die Hinzunahme der Gesamtnetzlänge als ein drittes Kriterium zur Bewertung der Entwürfe verdeutlichte sich ein Dilemma, das derjenige, der schon einmal die interaktive Verbesserung eines Floorplan mit mehreren Makrozellen versucht hat, sicher kennt: Versucht man, durch geschickte Formauswahl (im Sinne der Zuordnung von Subzell-Flächen zu Floorplan-Räumen) Verschnittflächen zu beseitigen, erhöhen sich fast zwangsläufig einige Netzlängen (und eventuell auch die Gesamtnetzlänge). Legt man stattdessen hohen Wert auf die Minimierung der Schnittnetze oder die Längen zeitkritischer Netze, muß man unter Umständen eine ungünstige Formauswahl in Kauf nehmen. Dieses ist eine Erklärungsmöglichkeit für die beobachteten Korrelationen zwischen Chipfläche und Gesamtnetzlänge und (der stärkeren) zwischen maximaler Verzögerungszeit und Gesamtnetzlänge.

Bei den Untersuchungen zur globalen Verdrahtung wurden Ergebnisse, die in [Fri98] nur angerissen waren, auf eine solide Basis gestellt. Ein Vergleich der drei Algorithmen pdGR, pdGR-N und pdGR-A zeigte, daß im Submicron-Bereich und vor allem im Deep-Sub-Micron-Bereich die netzlängenminimale Verdrahtung ohne Berücksichtigung des Verzögerungsverhaltens (pdGR-N) nicht konkurrenzfähig ist (vgl. Abb. 5.18). Die Anwendung der SERT-Heuristik auf jedes Netz ohne eine vorherige Timinganalyse zur Identifizierung der zeitkritischen Senken (pdGR-A) zeigte dagegen ein extrem schwankendes Verhalten: Manche Lösungen waren so gut wie die von pdGR, andere allerdings sogar schlechter als die von pdGR-N.

Schließlich wurden noch die durchschnittlichen Einspareffekte durch OTC-Verdrahtung (auf einer Hierarchieebene) anhand einer repräsentativen Auswahl der Beispiele und Technologien gemessen. Die Tabelle 6.2 zeigt die Zusammenfassung der Ergebnisse.

Tabelle 6.2: Durchschnittliche Einspareffekte bei OTC-Verdrahtung

Technologie	Beispiel	maxDel ohne	maxDel mit	%	area ohne	area mit	%
ES2_0.7	adr	14,02	13,87	1,1	98,2	91,3	7,0
SK2	lon.2	4,116	4,052	1,6	99,5	92,4	7,1
SK3	grp.2	8,89	8,67	2,5	95,4	87,5	8,3

Die durchschnittlichen Einsparungen bei der Gesamtnetzlänge fielen jeweils etwas höher aus als bei der Chipfläche: 8,2 % bei *adr* (ES2_0.7), 8,3 % bei *lon.2* (SK2) und 9,7 % bei *grp.2* (SK3). Die moderaten Einsparungen bei der Chipfläche haben ihre Ursache u. a. darin, daß das Ausgangspotential, also die Verdrahtungsflächen, die nicht aus Verschnittflächen resultieren, bei den Beispielen nicht allzu hoch war, wie ein Blick auf einige Flächenpläne beispielhaft verdeutlichte.

Schließlich läßt sich feststellen, daß mit dem *planner.v6* ein Tool zum timing-driven Floorplanning in das PLAYOUT-Entwurfssystem integriert wurde, das

- Chipfläche und maximale Verzögerungszeit gleichermaßen gut optimiert
- verschiedene Strategien der zeitgesteuerten Platzierung anbietet, die auf robusten Algorithmen mit intuitiv einstellbaren Parametern beruhen
- hierarchische Timinggraphen und die konzeptionelle Erweiterung von Shape Functions zu SD-Kurven für die Modellierung des Zeitverhaltens im Kontext des hierarchischen top-down Entwurfs verwendet
- standardmäßig einen Algorithmus zum performance-driven Global Routing verwendet, der von der Timinganalyse identifizierte zeitkritische Senken bevorzugt behandelt
- für die globale Verdrahtung das Modell eines Routinggraphen benutzt, das Kanalverdrahtung und OTC-Verdrahtung gleichermaßen unterstützt
- eine interaktive graphische Oberfläche und eine Experimentierumgebung für aktuelle und zukünftige Technologien bietet.

