

Sharing and Binding for General Circuits

Benedikt Lipinski
Interaktionstechnik und Design
Hochschule Hamm Lippstadt
Lippstadt, Germany
benedikt.lipinski@stud.hshl.de

Abstract

CONTENTS

I	Introduction	1
I-A	Problemstellung	1
I-B	Logische Bausteine	1
I-B1	AND-Gatter	1
I-B2	OR-Gatter	1
I-B3	Multiplexer	1
I-B4	FPGA	1
I-C	Strategien zur Architektur-Optimierung	1
I-D	Allocation	2
I-E	Binding	2
I-F	Sharing	3
II	Grundlegendes	3
II-A	Kompatibilitäts- und Konfliktgraphen	3
II-B	Resource Dominated circuits	4
III	General Circuits	4
III-A	Allgm.	4
III-B	Baugruppen	4
III-B1	Register	4
III-B2	Steuerlogik	5
III-B3	Verkabelung	5
III-B4	kontroll einheiten	5
IV	Sharing and Binding for General Circuits	5
IV-A	Unconstrained minimum - Area Binding	5
IV-B	Performance Constrained Binding	5
IV-C	Performance Directed Binding	5
V	Gleichnis,vorgestellter Algorithmen	5
VI	Ausblick	5
References		5
	//Löschen!!!!	

I. INTRODUCTION

Durch das menschliche Verlangen, Problemstellungen zugunsten schwindender Komplexität und erhöhten Komforts zu automatisieren und zu brechen, wurden genau diese Anforderungen auf diejenigen elektrischer Schaltungen umgelegt.

A. Problemstellung

Nicht erst durch moderne Entwicklungen wie digitale Vernetzung mit dem Internet der Dinge oder das autonome Fahren wird die Betrachtung zeitlicher und finanzieller Aspekte für die Entwicklung multidimensionaler Systeme notwendig. Allerdings lassen genau diese Entwicklungen die Komplexität sprunghaft ansteigen.

B. Logische Bausteine

Realisierbar werden elektrische Verschaltungen in der Digitaltechnik erst durch den Einsatz sogenannter Logikbausteine, die miteinander kombiniert ein vorher genau bestimmtes Verhalten der geplanten Schaltung verursachen. Hierbei übernehmen verschiedene Bausteine verschiedene Aufgaben, die mal weniger und mal mehr komplizierte Operationen beinhalten.

1) *AND-Gatter*: Eine der Grundoperationen, die in fast jeder Schaltung zu finden ist, ist die AND-Operation, zu deutsch die UND-Operation. Hierbei wird der Eingang erst auf logisch 1 geschaltet, wenn alle Eingänge dies ebenfalls sind.

x	y	Out
0	0	0
0	1	0
1	0	0
1	1	1

2) *OR-Gatter*: Das Ergebnis des OR-Gatters ist in dem Fall wahr, wenn einer seiner Eingänge wahr ist.

x	y	Out
0	0	0
0	1	1
1	0	1
1	1	1

Simple Schaltungen lassen sich durch das gezielte Aneinanderreihen von AND- und OR-Gattern gut realisieren, sollte aber

3) *Multiplexer*: Multiplexer werden genutzt, wenn eine parallele Verarbeitung aus unterschiedlichsten Gründen nicht möglich ist. Das kann zum einen eine nicht ausreichende Verfügbarkeit an Kanälen sein, aber auch die bewusste Entscheidung, weniger Kanäle aus Kostengründen zu benutzen, sein. Dieser Schritt kann gewählt werden, wenn Verläufe zwar parallel laufen, aber nur exklusiv ausgeführt werden. Das bedeutet, dass nach einer elementaren Entscheidung nur einer der beiden Stränge ausgeführt werden kann. Beeindruckend allerdings ist, dass Multiplexer trotz ihrer fortgeschrittenen Eigenschaften gegenüber der Grundbausteine doch recht simpel aus genau diesem erzeugt werden können. So ergibt sich für einen einfachen Multiplexer das Schaltbild aus zwei AND-Gattern, einem NOT-Gatter und einem OR-Gatter.

4) *FPGA*: Der Einsatz von diesen Grundelementen ist für das Erreichen des Ziels völlig hinreichend, doch benötigen sie eine sehr hohe Fläche, auf der sie verbaut werden müssen. Ein zu erreichendes Ziel ist, die Fläche auf der die Schaltung realisiert werden muss, zu verringern. Dies ist durch die Verwendung von Halbleitern und die Entwicklung von integrierten Schaltkreisen auch erfolgreich gelungen, allerdings besitzen diese sogenannten ICs den Nachteil, dass ihre Schaltung- wie die einer Realisierung mit Bausteinen- auch fest gebunden ist. Ein Nachteil, den FPGA (Field Programmable Gate ARRAY) Bauteile nicht besitzen, ihr Vorteil ist die Zusammensetzung aus zusammengeschalteten Baugruppen aus Logikbausteinen, FlipFlops und deren Verbindung mit Multiplexern. In FPGAs kann das Verhalten der Eingänge zu den Ausgängen über..... "programmiert" werden.

C. Strategien zur Architektur-Optimierung

Mit steigender Komplexität der Verschaltungen steigt zwingend auch die Anzahl der zu nutzenden Bauteile. Zudem steigt auch die Verwendung performanterer Bausteine stetig, um die Abarbeitung bezüglich Leistung und aller zeitlichen Deadlines einhalten zu können. Nachteilig ist vor dem Hintergrund ebenfalls, dass Bausteine mit einer höheren Leistung oftmals auch einen erhöhten Platzbedarf einfordern [2, S.327].

Ziel ist es, Strategien umzusetzen, die die Chipfläche verringern, ohne dabei die Leistungsfähigkeit der Schaltung zu beeinflussen. Leider ist dieses Problem nicht mit der Einhaltung beider Anforderungen zu vereinbaren, wodurch zum Beispiel die Verminderung der Kosten durch Senkung der Chip-Anzahl und Fläche immer auch zum Nachteil der Performance agiert. Der Zusammenhang zwischen Fläche und der erzeugten Leistungsverminderung ergibt sich aus der Fläche a in der Einheit Gatteräquivalent ($GateEquivalentGE$), das die Größe des simpelsten Gatters widerspiegelt [2, S.326] und der Performance, die den Kehrwert der Ausführungszeit T_{ges} widerspiegelt $\frac{1}{T_{ges}}$, die sich wieder aus der Taktzeit T_c multipliziert mit der Latenz λ zusammensetzt [2, S.325-326]. Eine mögliche Realisierung der Schaltung mit einer bestimmten Performance bei einer bestimmten Fläche lässt sich als Punkt in einem Diagramm darstellen, dessen Achsen sowohl die Chipfläche, die Latenz und den Takt abbilden [2, S.326]. Dieses Vorgehen [?] ist in diesem Fall notwendig, da mehr als eine Variable optimiert werden muss, und es dementsprechend mehr als eine gültige Lösung dieses Problems gibt. Als mögliches Hilfsmittel kann hier die Pareto-Optimierung verwendet werden. Die Pareto-Optimierung erzeugt eine Kompromisslösung durch die Gewichtung jeder Einzelfunktion. Die Idee dahinter ist, dass die Gewichtung die Wichtigkeit der Einzelfunktion repräsentiert und ihr bei steigendem Gewicht mehr Aufmerksamkeit verleiht [4, S.45]. Als ergebnis erhalten wir sogenannte Pareto punkte, diese punkte in unserem entwurfsraum stellen verschiedenste verhältnisse unserer optimierungs variablen dar. Eine Pareto optimale lösung, ist die lösung, die gültig gegenüber aller gestellten anforderungen ist und die das best mögliche ergeb-

nis aller variablen liefert, ohne das sich eine verschlechtert nur um eine andere weiter zu optimieren.

D. Allocation

Nicht nur, aber vor allem aus Kosten und Platzgründen stehen der Entwicklung selten unbegrenzte resources zur Verfügung. Aus diesem Grund kann die Betrachtung der zur Verfügung stehenden resources zwingend notwendig werden. Somit ist das ergebnis der allokierung nach Teich [1], eine funktion $\alpha(r_k)$ die angibt wie viele verfügbare Bauteile es zu einer Jeweiligen Ressourcentyp gibt.

Damit würde, wenn vom Ressourcentyp Multiplizier: r_1 , Beispielsweise 1 Einheiten zur Verfügung stehen die Allokierungsfunktion $\alpha(r_1) = 2$ lauten. Die Funktion der Allokation zeigt deshalb ziemlich genau welche probleme beziehungsweise zeitliche veränderungen entstehen durch die grenzen, die durch die begrenzte verfügbarkeit an Resources gegeben werden. Dies kann ein nützlicher zwischenschritt sein um den überblick zu behalten, den Teich mittels eines Ressourcen graphen anschaulich darstellt. In vielen Literaturen allerdings wird dieser schritt auch in die Bindung integriert und nicht extra behandelt.

E. Binding

Im unterschied zur Allokierung wird in der Bindung schließlich festgelegt welche Resource welche operation zugeteilt bekommt, also die Zuweisung von Operationen zu bestimmten hardware Komponenten [3]. in einer Utopischen vorstellung in der Kosten und Fläche keine rolle spielen würden, könnte für jeden Rechenschritt, jede Operation, also zum beispiel für jede Multiplikation oder jeden vergleich eine eigene Komponente Benutzt werden. Leider ist diese vorstellung in der Praxis nicht durchführbar, da jedes gesparte bauteil auf eine Große stückzahl gerechnet den Kostendruck lindert. Das ist zudem auch nicht zwingend nötig, da Beispielsweise ein Vergleich nach eine Addition, welche beide durch eine ALU realisiert werden können, nicht unbedingt durch 2 verschiedene Hardware einheiten umgesetzt werden müssen, da der vergleich erst stattfinden kann wenn das ergebnis der Addition vorliegt. genannt wird dies in diesem fall ein Optimal Resource sharing oder zu Deutsch ein Optimales ausnutzen einer Resource [2].

Die Bindung bietet ein Tool, das veranschaulicht wie optimal die vorhandene schaltung auf und mit den gegebenen Bauteilen realisiert wird. Da das engültige Ziel eine Optimale ausnutzung der vorhandenen resources ist müssen nun schrittweise die benötigten Bauteile minimiert werden, dies wird erreicht indem anfangs jede Operation einem bauteil zugeordnet wird, es wird ein sogenannte dedizierte Ressource erzeugt, ein spezialfall [3, S. 150]

Dem Bild ist zu entnehmen das in diesem beispiel für die zuordnung jeder Aufgabe zu einer Resource: 11 Bauteile benötigt werden. Die sich aus 3 Multiplikationen = r_1 und aus 8 operationen für eine Arithmetisch logisch einheit r_2 ergibt.

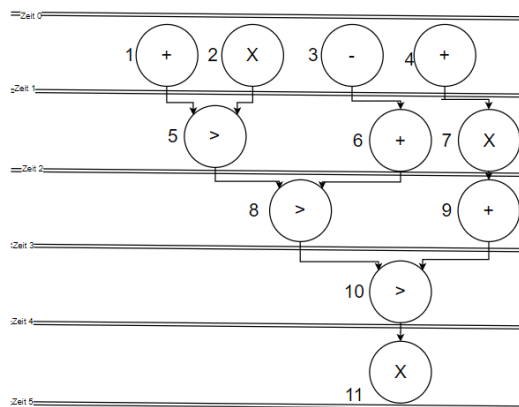


Fig. 1. Beispiel zur Bindung

v_i	$\beta(v_i)$	$\gamma(v_i)$
v_1	r_2	1
v_2	r_1	1
v_3	r_2	2
v_4	r_2	3
v_5	r_2	4
v_6	r_2	5
v_7	r_1	2
v_8	r_2	6
v_9	r_2	7
v_{10}	r_2	8
v_{11}	r_1	3

Da es sich, um ein recht einfach

zu überblickendes beispiel handelt, lässt sich durch einen geübten blick bereits erkennen wie eine Optimierung aussehen könnte. Im ersten Zeitslot ist abzulesen, das 3 Arithmetische operationen und eine Multiplikation durchgeführt werden, somit werden 3 ALU Einheiten und ein Multiplizierer benötigt. Im weiteren verlauf lässt sich unter anderem erkennen das in keinem weiteren zeitslot mehr als 3 ALU einheiten gebraucht werden, deswegen ist dies auch unsere minimal benötigte anzahl an Resources dieses typs. Ganz genau deutlich wird dies durch eine veranschaulichung in einer tabelle.

Zeitslot	ALU	Multiplizierer
1	3	1
2	2	1
3	2	0
4	1	0
5	0	0

Aus der Tabelle kann nun abgelesen werden, das 3 Resources des Typs ALU benötigt werden und 1 des Typs Multiplizierer. Übertragen auf die obere Tabelle entsteht nun folgendes Bindung.

v_i	$\beta(v_i)$	$\gamma(v_i)$
v_1	r_2	1
v_2	r_1	1
v_3	r_2	2
v_4	r_2	3
v_5	r_2	1
v_6	r_2	2
v_7	r_1	1
v_8	r_2	1
v_9	r_2	2
v_{10}	r_2	1
v_{11}	r_1	1

Somit wurden in diesem Beispiel 5 Alu

einheiten die einen Anteil von 62,5% ausmachen eingespart. Bei den Multiplizierern wurde die benötigten Ressourcen auf $\frac{1}{3}$ reduziert.

F. Sharing

das sharing beschreibt zusammen mit dem Binding, die effektive mehrfachnutzung der Ressourcen, welches der minimierung der verwendeten hardware dient und somit zu reduzierung von Flächen und kosten problemen beiträgt. wobei das ziel des Sharing die frage nach dem Algoritmus zur optimierung ist und anschließend mit einem Binding endet. In dem Kapitel Binding wurde ein einfach im Kopf zu lösendes Beispiel gezeigt, leider bleibt es in der Realität nicht immer bei so einfach zu lösenden systemen, so dass da schnell der überblick verloren gehen kann und zudem ein system unter umständen nicht zu lösen wäre. Aus diesem grund befasst sich dass sharing mit der frage wie ein vorhandenes system effektiv optimiert werden kann. Zu den Verwendeten Algorithmen zählen unter anderem aus dem bereich der Resource Dominated Circuits das Register sharing, Multiport Memory Binding und das Bus sharing and Binding und für den bereich der general circuits, sind es Unconstrained minimum - Area Binding, Performance Constrained Binding und zu letzt Performance Directed Binding.

II. GRUNDLEGENDES

A. Kompatibilitäts- und Konfliktgraphen

Als Voraussetzung, um überhaupt eine sinnvolle Bindung beziehungsweise Allokierung erzeugen zu können muss sich im 1. schritt einmal mit dem eigentlichen problem befasst werden. Um das Problem darstellen zu können wird ein Graph aufgestellt, der die eigenschaften der elemente beschreibt. Dieser Schritt ermöglicht die betrachtung der einzelnen operationen mit einander und stellt auf welche mit einander kompatibel sind und welche ein kritisches verhalten zu einander aufweisen.

Zwei oder mehr Operationen sind mit einander kompatibel und können an die Gleiche Resource gebunden werden wenn sie nicht mit einander konkurrieren und wenn sie dem gleichen Ressourcen-typen angehören [3, S.231]

Das bedeutet konkret, dass aufgaben mit einander kompatibel sind, wenn sie nicht im gleichen Zeitslot neben einander anliegen und wenn sie des gleichen typs sind zum Beispiel Addierer. Die entwicklung der ges graphensystems kann

grundsätzlich über 2 ansätze erfolgen, zum einen kann ein Kompatibilitätsgraph oder Verträglichkeitsgraph entwickelt werden, dieser graph gibt an welche Operationen verträglich zu einander sind und somit keine probleme darstellen komplementär dazu lässt sich durch den Konflikt graphen angeben welche Operationen im konflikt zu einander stehen und eine genauere betrachtung bedürfen.

Ein Ressourcen Verträglichkeitsgraph ist definiert durch $G_+(V, E)$ in denen die Knoten durch $V = \{v_i; i = 1, 2, \dots, n\}$ definiert werden und die Kanten durch $E = \{\{v_i, v_j\}; i, j = 1, 2, \dots, n\}$. Die Knoten stellen die Operationen an sich dar, wohingegen die Kanten die Kompatiblen Operationspaare angeben [3, S.231].

Zusammen gefasst nach Teich [1, S. 181-182] lassen sich folgende verträglichkeiten für den kompatibilitätsgraphen aufführen:

demnach ist eine **Schwache verträglichkeit** gegeben, wenn zwei knoten des Graphen eines Graphen mit einander verbunden sind und des gleichen Types angehören also zum beispiel eine Addition und ein Vergleich, da beide durch eine Arithmetisch logische einheit Verarbeitet werden können. [?]

Ablaufplan verträglich ist ein graph hingegen, wenn seine Knoten als erste bedingung vom gleichen typ sind, also wie bei der schwachen verträglichkeit zum Beispiel aus 2 operationen vom typ Multiplizierer. Zum zweiten muss die bedingung erfüllt sein, dass die startzeit des ersten knoten vor der des zweiten knotens liegen muss. Das bedeutet, dass Knoten die Ablaufverträglich sind nicht in dem selben zeitslot ausgeführt werden dürfen, andernfalls wären sie nicht Ablaufplan verträglich [?].

Als letztes definiert sich die **Starke verträglichkeit**, durch zwei Knoten die wieder vom gleichen typ sind und zudem von einandern abhängig sind. Das bedeutet die beiden knoten sind in einem logischen strang durch einen pfad im ablaufplan mit einander verbunden. [?]

Die Graphische umsetzung des Verträglichkeitsgraphes mit dem Beispiel aus den vorherigen kapiteln sieht dann wie folgt aus:

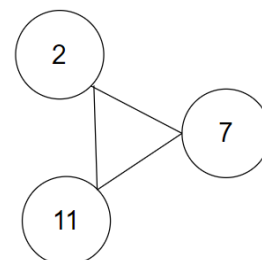


Fig. 2. Verträglichkeitsgraph Multiplizierer

Aus den vorherigen kapiteln, wissen wir bereits, dass für die Optimierung der verschaltung nur ein Multiplizierer benötigt wird und genau das können wir in diesem beispil nun auch

grafisch betrachten. Zu sehen ist, dass jeder Knoten mit einer kante verbunden ist und somit sind alle Operationen mit einander Ablaufplan verträglich, weswegen der einsatz von nur einem Multiplizierer nun auch Optisch besser nachzuvollziehen ist.

Als nächsten schritt werden nun sogenannte Cliques eingeze-

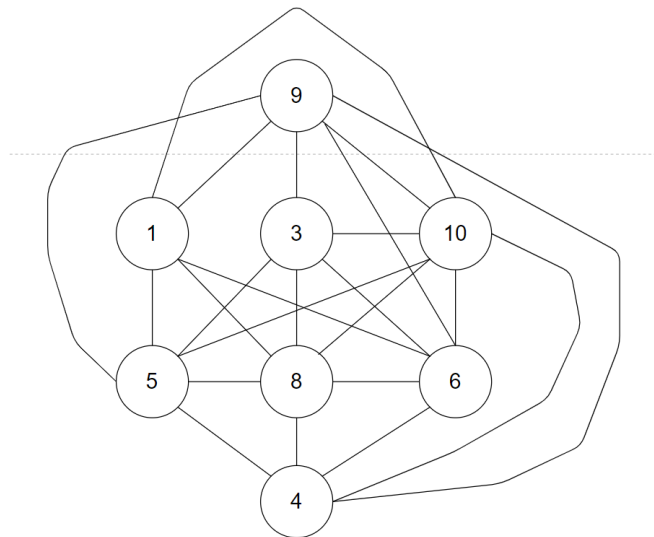


Fig. 3. Verträglichkeitsgraph Arithmetisch logische einheit mit Cliques

ichnet, sie repräsentieren die zusammenfassung Mehrerer Operationen und zeigen ihre zusammenführung auf eine Komponente, so das für jede clique genau eine Komponente eingesetzt wird. Zu sehen ist, das Bei dem Verträglichkeitsgraphen in diesem Beispiel für die Multilplizierer genau eine Clique eingezeichnet werden muss. Da mit steigender Komplexität der aufgabe, auch die Komplexität für das einzeichnen der Cliques steigt, bietet sich dieses verfahren leider nur für die anwendung an simple verschaltungen an, der Grund hierfür ist die klassifizierung dieses problems als NP-Hartes Problem. [2, S.361]

B. Resource Dominated circuits

Das "Sharing and Binding for Resource Dominated Circuits" (Verteilung und Bindung für Ressourcen dominierte Schaltkreise) Bedeutet, im folgenden die umsetzung und kombinierung, des Wissens und der Grundlagen aus den voran gegangen Abschnitten. Zur Erinnerung ist, das gesuchte ergebnis eine Bindung, das bedeutet eine zuordnung von tatsächlicher hardware und Operationen in denen eine maximale menge an Operationen einer minimalen menge an Hardware zugeordnet wird. hierfür ist im vorfeld zwingen festzustellen ob nicht druch andere dynamische umwelteinflüssen und anforderungen, eine grundsetzliche begrenzung der zur verfügung stehenden harware und fläche besteht (allokierung). Desweiteren, um überhaupt feststellen zu können, welche operationen sich auf der selben hardware kombinieren lassen, muss ein Kompatibilitäts-/Konflikt-graph aufgestellt

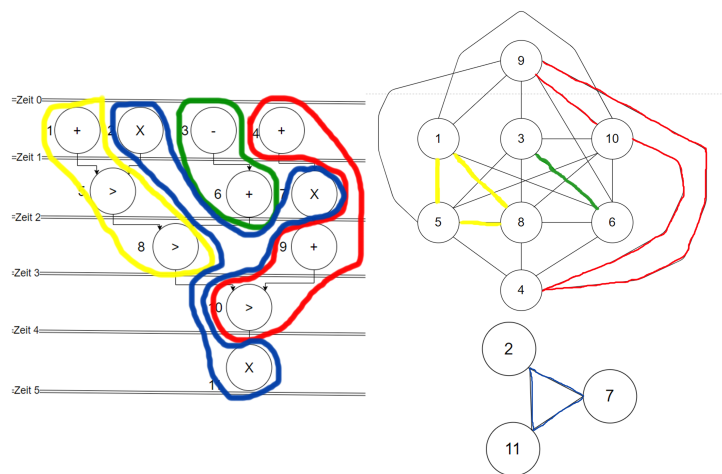


Fig. 4. Ablaufplan und verträglichkeitsgraphen mit Cliquesbildung

werden, mit dessen ergebnis sich operationen zu Cliques bündeln lassen. Leider ist die Lösung des vorliegenden problems nicht derart einfach, als einziges ziel lediglich die fläche und somit den preis der verschaltung zu optimieren, gerade in Echtzeitsystemen muss auch immer die aufmerksamkeit zu gleichen teilen auf der einhaltung der zeitlichen vorgaben bleiben. Deswegen betrachtet das sharing and Binding für Resource Dominated Circuits, die Optimierung der Schaltung, bezüglich Fläche(Kosten) und Latenz(Zeitschranken), mittels der Optimierung der wirkenden Schaltung mittels vermin derung der eingesetzten Hardware und doppelnutzung dieser. [3, S. 156]

III. GENERAL CIRCUITS

Die betrachtung einer Schaltung unter dem aspekt der RESource Dominated Circuits ähnelt allerdings eher der betrachtung einer Idealen Komponente, als einer in denen auch störgrößen und andere umwelteinflüsse einer gewichteten zuwendung erfahren.

A. Allgm.

Aus diesem Grund wird für das Sharing and Binding für General Circuits (Verteilung und Bindung für Generelle schaltkreise) nicht nur latenz und fläche der Operierenden Elektronik betrachtet, sondern auch derer, die steuernde und speichernde eigenschaften besitzen, zudem wird auch die verkabelung dieser nicht außen vor gelassen. [3, S. 156]

B. Baugruppen

1) Register: Da in schaltungen, wie sie in allen modernen systemen vorhanden sind, die über eine einfache input/output addition hinaus gehen Daten und werte über einen einzelnen zyklus hinaus gespeichert werden müssen, kommen in vielen schaltungen speicehrelemente zum einsatz beispielsweise müssen in Reglungs-Schaltungen werte gespeichert werden, die über mehrere zyklen zur verfügung stehen. // Grundlegend

verhält sich die bindung und das Teilen von registern ganz ähnlich zu den vorgängen bei normalen operatoren. Auch sind die vorgaben, bei denen REgister geteilt werden können gleich zu denen der Operatoren, das bedeutet Register können geteilt werden, wenn sie nicht in konkurrenz zu einander stehen, das bedeutet wenn sie exklusiv zu einander stehen oder erst nach einander auftreten. Nicht analog allerdings ist hier die definition der Lebenszeit einer variable und dementsprechend auch der register nutzung ist. Die lebenszeit ergibt sich nach De Micheli aus dem geburtszeitpunkt (*former*) und dem Todeszeitpunkt (*latter*) einer variable. Der Geburtszeitpunkt ist der Zeitpunkt der erstellung durch das ergebnis einer Operation und der todeszeitpunkt ist als der späteste zeitpunkt anzunehmen in dem die Variable als input referenziert wird [3, S. 240]

2) *Steuerlogik:*

3) *Verkabelung:*

4) *kontroll einheiten:*

IV. SHARING AND BINDING FOR GENERAL CIRCUITS

A. *Unconstrained minimum - Area Binding*

B. *Performance Constrained Binding*

C. *Performance Directed Binding*

V. GLEICHNIS,VORGESTELLTER ALGORITHMEN

VI. AUSBLICK

REFERENCES

- [1] Jürgen Teich, Christian Haubelt, Digitale Hardware/Software-Systeme, Synthese und Optimierung, 2. Auflage, Springer, 2007
- [2] Dr. Walter Lange, Prof. Dr. Martin Bogdan, Entwurf und Synthese von Eingebetteten Systemen, Ein Lehrbuch, Oldenbourg Verlag München, 2013
- [3] DEMICH
- [4] Martin Pieper, Mathematische Optimierung, Eine Einführung in die Kontinuierliche Optimierung mit Beispielen, Springer Spektrum, 2017