Small Logic-based Multipliers with Incomplete Sub-Multipliers for FPGAs

Andreas Böttcher
Martin Kumm

31st Symposium on Computer Arithmetic (ARITH)
June 10-12, 2024 Malaga, Spain

HOCHSCHULE FULDA
UNIVERSITY OF APPLIED SCIENCES

Agenda

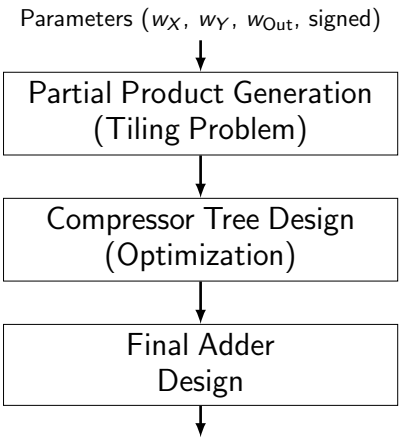
- Previous work on multiplier tiling
- Proposed incomplete tiles
- Results

Background / Motivation

- Multiplication is an elementary operation
- Research since 1950's
- Particularities of FPGAs
- Goals
  - reduce resource utilization
  - minimize critical path
  - minimize latency
  - more energy efficient
  - reduced (monetary) cost

Multiplier Design

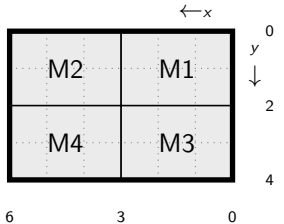
- 3-step approach
  - not independent
  - combined optimization



Multiplier Tiling

$$X \times Y = \left(x_h 2^{W_X} + x_l\right)\left(y_h 2^{W_Y} + y_l\right)$$
$$= \underbrace{x_h y_h}_{M_4} 2^{W_X+W_Y} + \underbrace{x_h y_l}_{M_3} 2^{W_X} + \underbrace{x_l y_h}_{M_2} 2^{W_Y} + \underbrace{x_l y_l}_{M_1}$$

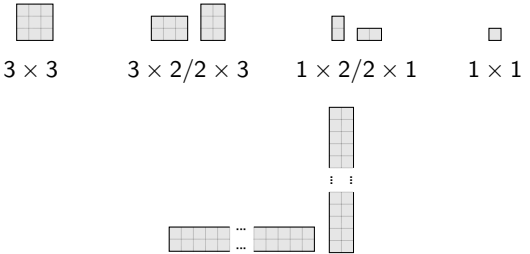
- Multiplier Tiling [1]
  - Assembling Multiplier
- Submultipliers
  - DSP-Units
  - LUT-Multipliers
- Optimization Problem



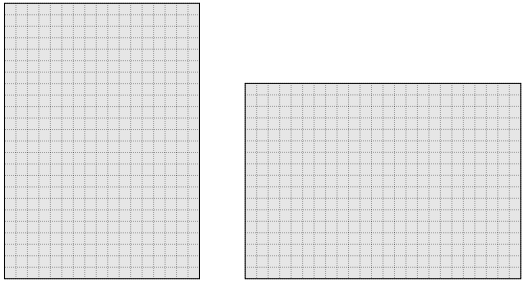
Tiling of a 6×4-multiplier

[1] S. Banescu, F. de Dinechin, B. Pasca: Multipliers for Floating-Point Double Precision and beyond on FPGAs, SIGARCH Comp. Arch. News, Sep.2010

Multiplier Tiles (for AMD 6,7,Ultrascale+,etc. Series)

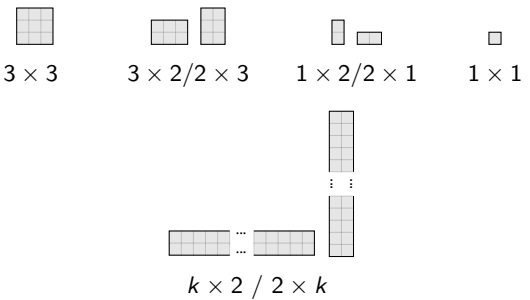


Geometric shapes of the LUT-based tiles



embedded DSP units

Multiplier Tile Properties (for AMD 6,7,Ultrascale+,etc. Series)



Geometric shapes of the LUT-based tiles

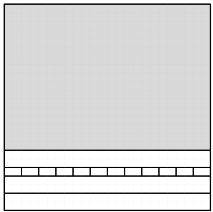
Properties of previous LUT- and DSP-based multipliers tiles [1]

Shape	A_t	$\text{cost}_t^{\text{tile}}$	E_t
1×1	1	1.65	0.625
1×2 / 2×1	2	2.3	0.87
2×3 / 3×2	6	6.25	0.96
3×3	9	8.9	1.011
$2\times k$ / $k\times 2$	$2k$	$1.65k + 2.3$	$\frac{2k}{1.65k+2.3}$
24×17 / 24×17	408	26.65	15.30

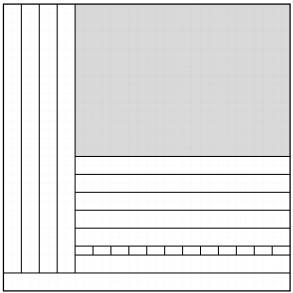
$$E_t = \frac{A_t}{\text{cost}_t^{\text{tile}}}$$

[1] M. Kumm, J. Kappauf, M. Istoan, P. Zipf, "Resource optimal design of large multipliers for FPGAs," in Symposium on Computer Arithmetic (ARITH), 2017

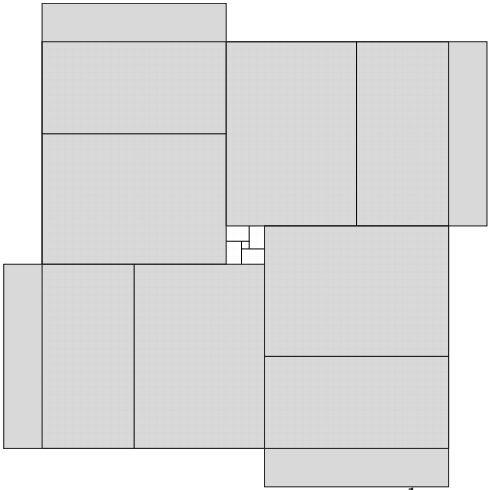
Multiplier Tiling Examples (Optimal)



Tiling of a 24×24 Multiplier [1]



Tiling of a 32×32 Multiplier [1]

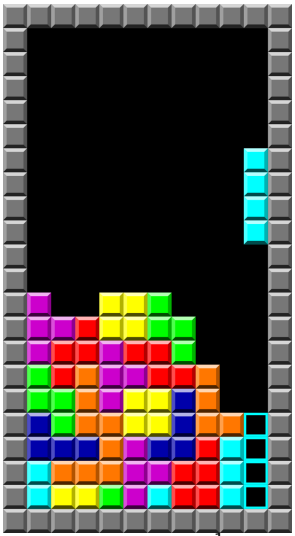


Tiling of a 53×53 Multiplier [1]

[1] M. Kumm, J. Kappauf, M. Istoan, P. Zipf, "Resource optimal design of large multipliers for FPGAs," in Symposium on Computer Arithmetic (ARITH), 2017

Multiplier Tiling

- Do you remember Tetris...
  - Filling area without gaps
  - Similar rules



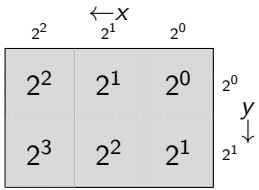
Tetris Game [1]

[1] https://upload.wikimedia.org/wikipedia/commons/archive/9/9c/20200827095319%21Typical_Tetris_Game.svg

Agenda

- Previous work on multiplier tiling
- Proposed incomplete tiles
- Results

Motivational Example



conventional 3×2-multiplier tile

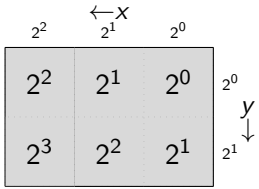
$P_{max} = (2^3 - 1) \times (2^2 - 1) = 21$

$\rightarrow$ 5bits $\rightarrow 5 \times$ 5LUTs $\rightarrow 3 \times$ 6LUTs

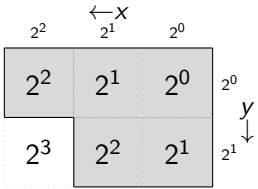
$\text{cost}_t^{\text{tile}} = 3 + 5 \times 0.65 \frac{\text{LUT}}{\text{bit}} = 6.25\text{LUT}$

$E_t = \frac{A_t}{\text{cost}_t^{\text{tile}}} = \frac{6}{6.25} = 0.96$

Motivational Example



conventional 3×2-multiplier tile



3×2-tile with omitted position

$P_{max} = (2^3 - 1) \times (2^2 - 1) = 21$

$\rightarrow$ 5bits $\rightarrow 5 \times$ 5LUTs $\rightarrow 3 \times$ 6LUTs

$\text{cost}_t^{\text{tile}} = 3 + 5 \times 0.65 \frac{\text{LUT}}{\text{bit}} = 6.25\text{LUT}$

$E_t = \frac{A_t}{\text{cost}_t^{\text{tile}}} = \frac{6}{6.25} = 0.96$

$P_{max} = 21 - 8 = 13$

$\rightarrow$ 4bits $\rightarrow 4 \times$ 5LUTs $\rightarrow 2 \times$ 6LUTs

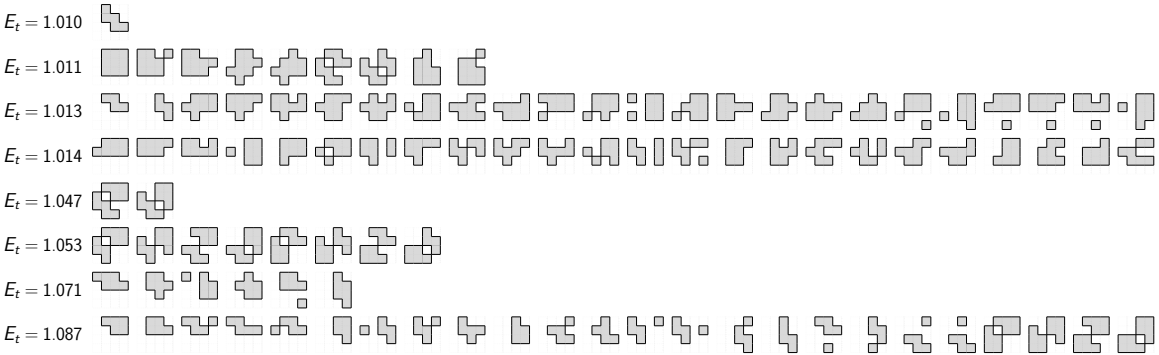
$\text{cost}_t^{\text{tile}} = 2 + 4 \times 0.65 \frac{\text{LUT}}{\text{bit}} = 4.6\text{LUT}$

$E_t = \frac{5}{4.6} = 1.087$

Search Space Exploration

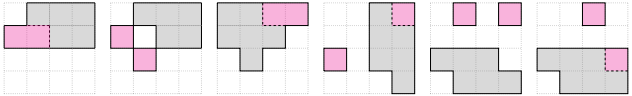
1. Generation of all possible tiles in 4×4
2. Tabulation of truth table
3. Logic minimization (Quine McClusky)
4. Recording of tiles according to efficiency

Identified Structures

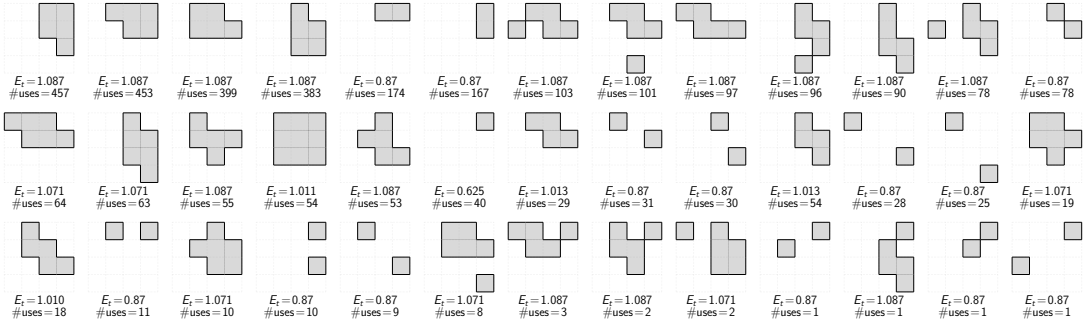


First 24 Tiles from the efficiency classes above E_t =1.0 of the 4×4 search space

Regularization



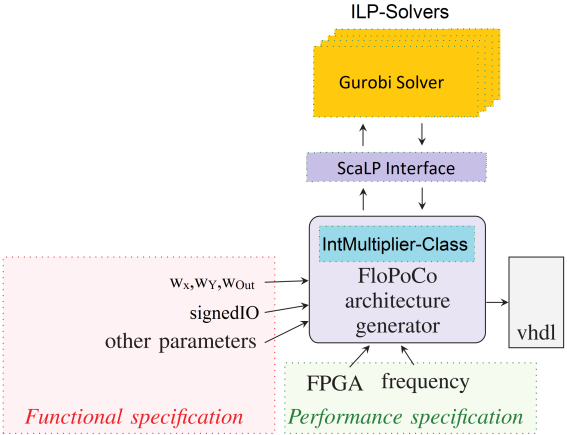
Helper tile decomposition



Geometric shapes of incomplete tiles used for experiments

Implementation of the Tiling Process

- Integration in FloPoCo (flopoco.org)
- Tiling & compression by previous ILP formulation [1]
- Solving by Gurobi
- Procession of the solution
- Instantiation of the individual multipliers
- Compressor tree synthesis
- VHDL generation



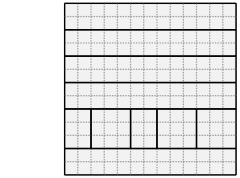
Realization of the Multiplier Design

[1] A. Böttcher and M. Kumm, "Towards globally optimal design of multipliers for FPGAs," IEEE Transactions on Computers, vol. 72, pp. 1261–1273, 2023

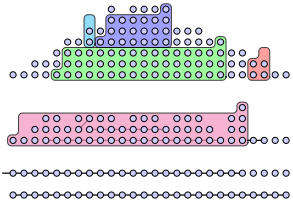
Agenda

- Previous work on multiplier tiling
- Proposed incomplete tiles
- Results

Results - Comparison to Previous Tiling

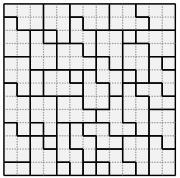


Tiling (91LUT)

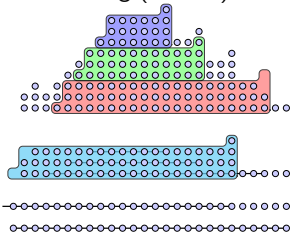


Compression (48LUT)

Total 91+48=139



Tiling (70LUT)



Compression (60LUT)

Total 70+60=130

Tiling and compression of a 13×13-multiplier

Results - Comparison to Previous Tiling

LUT results and (improvement) compared to previous tiling [1] for $W_X \times W_Y$ multipliers

$W_Y \backslash W_X$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	**1 (0%)**	1 (0%)	2 (0%)	2 (0%)	3 (0%)	3 (0%)	4 (0%)	4 (0%)	5 (0%)	5 (0%)	6 (0%)	6 (0%)	7 (0%)	7 (0%)	8 (0%)	8 (0%)
2	1 (0%)	**2 (0%)**	3 (0%)	5 (0%)	6 (0%)	7 (0%)	8 (0%)	9 (0%)	10 (0%)	11 (0%)	12 (0%)	13 (0%)	14 (0%)	15 (0%)	16 (0%)	17 (0%)
3	2 (0%)	3 (0%)	**5 (0%)**	9 (10.0%)	12 (0%)	14 (6.6%)	16 (5.8%)	18 (5.2%)	21 (0%)	23 (4.1%)	25 (7.4%)	28 (0%)	30 (3.2%)	32 (3.0%)	34 (2.8%)	35 (7.8%)
4	2 (0%)	5 (0%)	9 (10.0%)	**12 (7.6%)**	15 (11.7%)	18 (5.2%)	20 (9.0%)	23 (11.5%)	26 (10.3%)	28 (12.5%)	31 (11.4%)	34 (10.5%)	37 (9.7%)	39 (11.3%)	42 (10.6%)	45 (10.0%)
5	3 (0%)	6 (0%)	12 (0%)	14 (17.6%)	**18 (10.0%)**	23 (-4.5%)	27 (3.5%)	30 (3.2%)	35 (0%)	38 (2.5%)	41 (2.3%)	-	49 (0%)	52 (1.8%)	56 (0%)	60 (0%)
6	3 (0%)	7 (0%)	14 (6.6%)	18 (-5.8%)	22 (0%)	**27 (0%)**	34 (0%)	38 (0%)	42 (0%)	46 (0%)	50 (0%)	54 (0%)	58 (0%)	62 (0%)	66 (0%)	70 (0%)
7	4 (0%)	8 (0%)	16 (5.8%)	20 (9.0%)	27 (3.5%)	34 (0%)	**37 (9.7%)**	42 (10.6%)	47 (7.8%)	52 (8.7%)	58 (6.4%)	63 (7.3%)	68 (8.1%)	73 (7.5%)	78 (8.2%)	83 (6.7%)
8	4 (0%)	9 (0%)	17 (10.5%)	-	30 (3.2%)	38 (0%)	42 (10.6%)	**48 (7.6%)**	54 (8.4%)	61 (6.1%)	67 (5.6%)	73 (5.1%)	79 (3.6%)	85 (4.4%)	91 (5.2%)	97 (3.0%)
9	5 (0%)	10 (0%)	21 (0%)	26 (10.3%)	36 (-2.8%)	42 (0%)	47 (6.0%)	54 (10.0%)	**61 (7.5%)**	69 (5.4%)	75 (6.2%)	81 (5.8%)	88 (4.3%)	96 (4.0%)	101 (4.7%)	108 (5.2%)
10	5 (0%)	11 (0%)	23 (4.1%)	28 (12.5%)	38 (2.5%)	46 (0%)	52 (10.3%)	61 (6.1%)	68 (6.8%)	**75 (8.5%)**	84 (5.6%)	91 (5.2%)	99 (3.8%)	106 (3.6%)	115 (1.7%)	121 (2.4%)
11	6 (0%)	12 (0%)	25 (3.8%)	31 (11.4%)	41 (2.3%)	50 (0%)	58 (6.4%)	68 (4.2%)	76 (5.0%)	83 (6.7%)	**93 (7.9%)**	102 (6.4%)	109 (7.6%)	117 (7.1%)	126 (5.2%)	134 (5.6%)
12	6 (0%)	13 (0%)	26 (7.1%)	34 (10.5%)	47 (-2.1%)	54 (0%)	65 (5.7%)	73 (6.4%)	81 (5.8%)	91 (5.2%)	102 (5.5%)	**110 (6.7%)**	120 (4.0%)	128 (3.7%)	137 (4.1%)	146 (4.5%)
13	7 (0%)	14 (0%)	30 (3.2%)	37 (9.7%)	49 (2.0%)	58 (0%)	68 (8.1%)	79 (3.6%)	88 (5.3%)	98 (4.8%)	109 (6.8%)	120 (6.2%)	**129 (6.5%)**	140 (4.1%)	149 (5.0%)	160 (3.6%)
14	7 (0%)	15 (0%)	32 (3.0%)	39 (11.3%)	52 (1.8%)	62 (0%)	73 (8.7%)	85 (4.4%)	94 (6.0%)	106 (3.6%)	118 (4.8%)	128 (5.1%)	139 (4.7%)	**153 (3.1%)**	162 (3.5%)	173 (2.8%)
15	8 (0%)	16 (0%)	34 (2.8%)	42 (10.6%)	56 (1.7%)	66 (0%)	78 (7.1%)	92 (3.1%)	101 (4.7%)	114 (2.5%)	126 (5.9%)	137 (4.1%)	149 (5.0%)	163 (2.9%)	**175 (5.4%)**	186 (3.6%)
16	8 (0%)	17 (0%)	35 (7.8%)	45 (10.0%)	60 (0%)	70 (0%)	84 (6.6%)	97 (3.0%)	108 (4.4%)	121 (2.4%)	135 (4.2%)	146 (3.9%)	159 (4.7%)	174 (1.6%)	186 (4.1%)	**199 (2.9%)**

[1] A. Böttcher and M. Kumm, "Towards globally optimal design of multipliers for FPGAs," IEEE Transactions on Computers, vol. 72, pp. 1261–1273, 2023

Results - Comparison to Previous Methods

LUT comparison to previous rectangular tiling, fractal synthesis ported to AMD and Booth array

Size	**proposed tiling** [LUT]	**rectangular tiling** [1] [LUT]	**fractal synthesis** [2] [LUT]	**Booth-Array** [3] [LUT]
3×3	5	5	6	9
4×4	12	13	n.a.	17
5×5	18	20	20	23
6×6	27	27	32	35
7×7	37	41	42	39
8×8	48	52	n.a.	51

[1] A. Böttcher and M. Kumm, "Towards globally optimal design of multipliers for FPGAs," IEEE Transactions on Computers, vol. 72, pp. 1261–1273, 2023

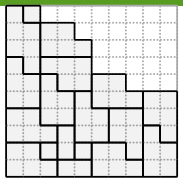
[2] M. Langhammer and G. Baeckler, "High density and performance multiplication for FPGA," in IEEE Symposium on Computer Arithmetic (ARITH), 2018

[3] M. Kumm, S. Abbas, and P. Zipf, "An efficient softcore multiplier architecture for Xilinx FPGAs," in Symposium on Computer Arithmetic (ARITH), 2015

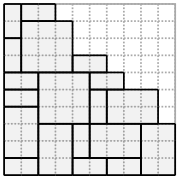
Results - Comparison to Previous Truncated Tiling

LUT results for truncated multipliers for $W = W_y = W_y = W_{\text{out}}$

W	proposed tiling	rectangular tiling [1]	improvement
1	1	1	0%
2	1	1	0%
3	3	7	57.1%
4	10	9	-11.1%
5	15	17	11.7%
6	22	22	0%
7	28	32	12.5%
8	37	40	7.5%
9	48	51	5.8%
10	57	67	14.9%
11	68	78	12.8%
12	80	89	10.1%
13	96	104	7.6%
14	111	121	8.2%
15	127	130	2.3%
16	144	148	2.7%



(a)



(b)

Tiling of a truncated 10×10-multiplier with (a) incomplete and (b) rectangular sub-multipliers

[1] A. Böttcher and M. Kumm, "Towards globally optimal design of multipliers for FPGAs," IEEE Transactions on Computers, vol. 72, pp. 1261–1273, 2023

Results - Packing Experiment

Packing experiment results for a 7×7 multiplier with various implementations

	Type	single mult.		#mult/FPGA		Utilization [%]	
		#LUTs	CPD [ns]	theory	actual	Slice	LUT
combinatorial	Proposed Tiling	36	4.5	1138	**1024**	**100.0**	89.9
	Rectangular Tiling [1]	42	4.8	976	964	**100.0**	**97.6**
	Xilinx IP speed opt. v12	58	**3.7**	706	559	98.6	79.0
	Xilinx IP area opt. v12	79	4.2	518	471	**100.0**	90.8
	Inferred multiplier	51	3.8	803	713	99.9	87.9
	Booth-Array [2]	39	3.9	1051	841	98.5	80.0
	Booth-Array [3]	**32**	3.8	**1281**	921	98.8	71.9
	Fractal Synthesis [4]	38	**3.7**	1078	822	99.6	76.1
pipelined	Proposed Tiling	36	3.0	1138	**1024**	**100.0**	**89.9**
	Rectangular Tiling [1]	41	3.2	1000	773	**100.0**	77.3
	Xilinx IP speed opt. v12	58	3.3	706	560	**100.0**	79.2
	Xilinx IP area opt. v12	77	3.2	532	422	99.9	79.3
	Inferred multiplier	51	3.6	803	713	**100.0**	87.8
	Booth-Array [3]	**32**	2.6	**1281**	919	98.9	71.7
	Fractal Synthesis [4]	38	**2.5**	1078	819	99.4	75.9

[1] A. Böttcher and M. Kumm, "Towards globally optimal design of multipliers for FPGAs," IEEE Transactions on Computers, vol. 72, pp. 1261–1273, 2023

[2] M. Kumm, S. Abbas, and P. Zipf, "An efficient softcore multiplier architecture for Xilinx FPGAs," in Symposium on Computer Arithmetic (ARITH), 2015

[3] E. G. Walters, "Partial-product generation and addition for multiplication in FPGAs with 6-input LUTs," in Asilomar Conference on Signals, Systems and Computers, 2014

[4] M. Langhammer and G. Baeckler, "High density and performance multiplication for FPGA," in IEEE Symposium on Computer Arithmetic (ARITH), 2018

Conclusion

- New category of multiplier tiles
  - Integration in previous tiling methods
  - Tiles offer advantages in practical cases
- Outlook
  - Further regularization
  - Generalized Optimization Model
  - Better (more precise) pipelining

The End

Thank you for your attention!