



Cramming CMACs into DSPs

Norbert Abel, Will Kamp
High Performance Computing Research Lab
Auckland University of Technology, New Zealand

Gianni Comoretto
Arcetri Astrophysical Observatory, Italy

15th February 2019 - C4SKA @ AUT

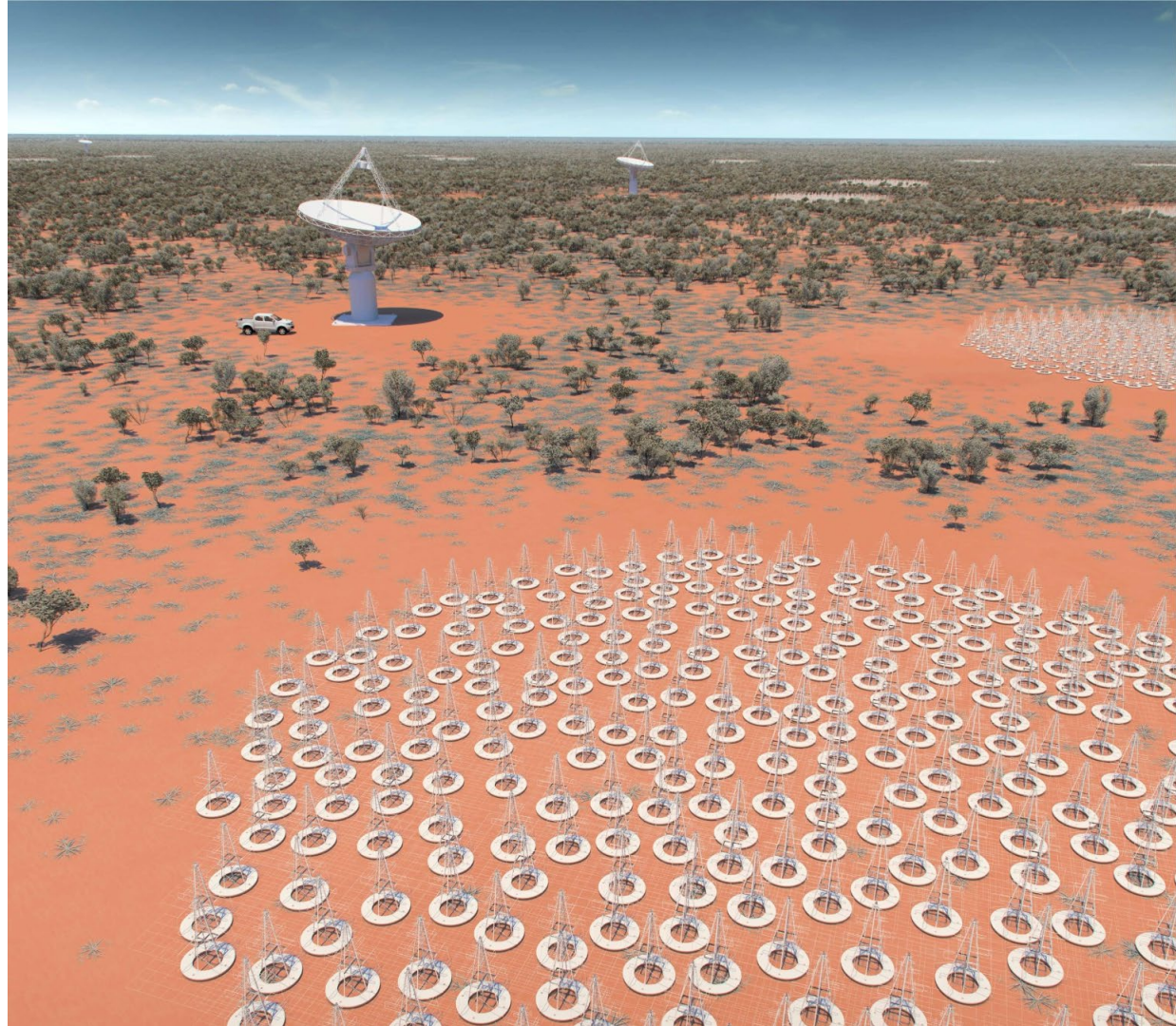
Published as “Complex Multiply Accumulate Cells for the Square Kilometre Array Correlators” at the ReConfig 2018 in Cancun, Mexico
Won “Best Paper Award”



SKA1 - Low

512 stations
(dual polarisation)

256 x 40 Gbps optical fibre links
back to the Central Signal
Processor (CSP)



CSP – Central Signal Processor



Cross-multiply every station with every other station: $O(N^2)$ operations.

Where $N = 2 \times 512 = 1024$ for Low.

Then accumulate to “beat down the noise” by the Central Limit Theorem.

Largely implemented with FPGA clusters: 288 Xilinx Ultrascale+ VU37P

SKA.Low = 288 FPGA x 1056 **CMAC** units @ 533 MHz



Complex Multiplication



Cartesian Multiplication:

$$\begin{aligned}z &= x \cdot y \\ &= (a + ib) \cdot (c + id) \\ &= (a \cdot c - b \cdot d) + i(a \cdot d + b \cdot c)\end{aligned}$$

4 Multiplies, 2 Additions/Subtractions

Karatsuba (Gaussian) Multiplication:

$$\begin{aligned}K_1 &= a \cdot c \\ K_2 &= b \cdot d \\ K_3 &= (a + b) \cdot (c + d) = (a \cdot c + b \cdot c + a \cdot d + b \cdot d) \\ z &= (K_1 - K_2) + i(K_3 - K_1 - K_2)\end{aligned}$$

3 Multiplies, 5 Additions/Subtractions

Complex Multiplication

Cartesian Multiplication:

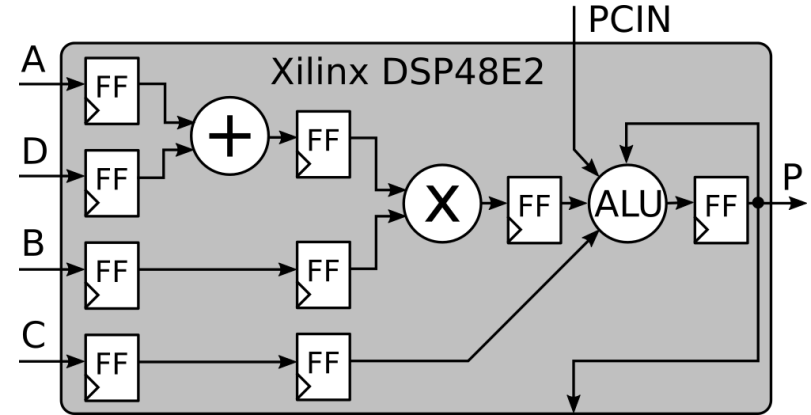
$$\begin{aligned}z &= x \cdot y \\ &= (a + ib) \cdot (c + id) \\ &= (a \cdot c - b \cdot d) + i(a \cdot d + b \cdot c)\end{aligned}$$

4 Multiplies, 2 Additions/Subtractions

Karatsuba (Gaussian) Multiplication:

$$\begin{aligned}K_1 &= a \cdot c \\ K_2 &= b \cdot d \\ K_3 &= (a + b) \cdot (c + d) = (a \cdot c + b \cdot c + a \cdot d + b \cdot d) \\ z &= (K_1 - K_2) + i(K_3 - K_1 - K_2)\end{aligned}$$

3 Multiplies, 5 Additions/Subtractions



Xilinx Ultrascale+
One 18x27b signed multiply.
=> 3168 DSPs

Complex Multiplication

Cartesian Multiplication:

$$\begin{aligned}z &= x \cdot y \\ &= (a + ib) \cdot (c + id) \\ &= (a \cdot c - b \cdot d) + i(a \cdot d + b \cdot c)\end{aligned}$$

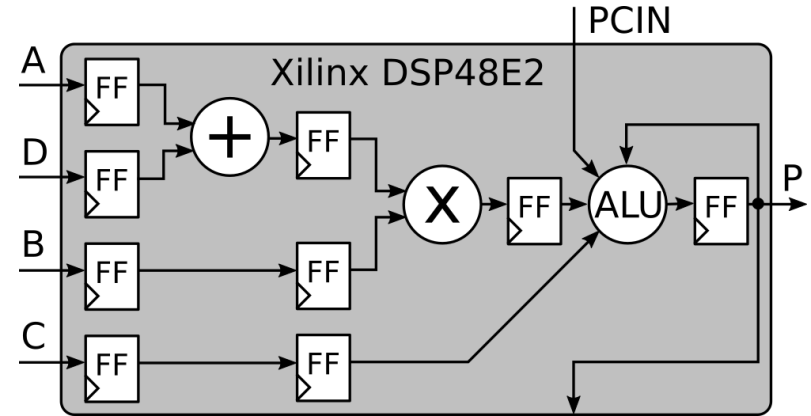
4 Multiplies, 2 Additions/Subtractions

Karatsuba (Gaussian) Multiplication:

$$\begin{aligned}K_1 &= a \cdot c \\ K_2 &= b \cdot d \\ K_3 &= (a + b) \cdot (c + d) = (a \cdot c + b \cdot c + a \cdot d + b \cdot d) \\ z &= (K_1 - K_2) + i(K_3 - K_1 - K_2)\end{aligned}$$

3 Multiplies, 5 Additions/Subtractions

Actually, we only need a $8b \cdot 8b$ multiplier!



Xilinx Ultrascale+
One 18x27b signed multiply.
=> 3168 DSPs

The cramming begins...



$$x = a + ib \Rightarrow x' = a + 2^w b$$

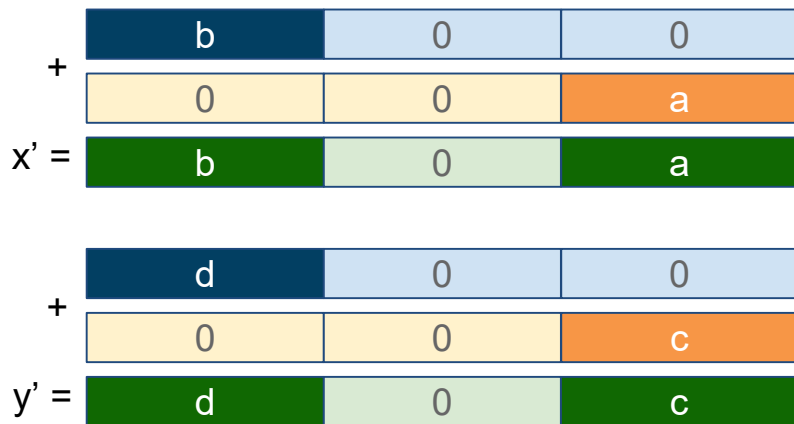
$$y = c + id \Rightarrow y' = c + 2^w d$$

$$z' = x' y'$$

$$z' = (a + 2^w b) \cdot (c + 2^w d)$$

$$z' = 2^{2w} b \cdot d + 2^{2w} b \cdot c + 2^w a \cdot d + a c$$

$$z' = 2^{2w} b \cdot d + 2^w (b \cdot c + a d) + a c$$



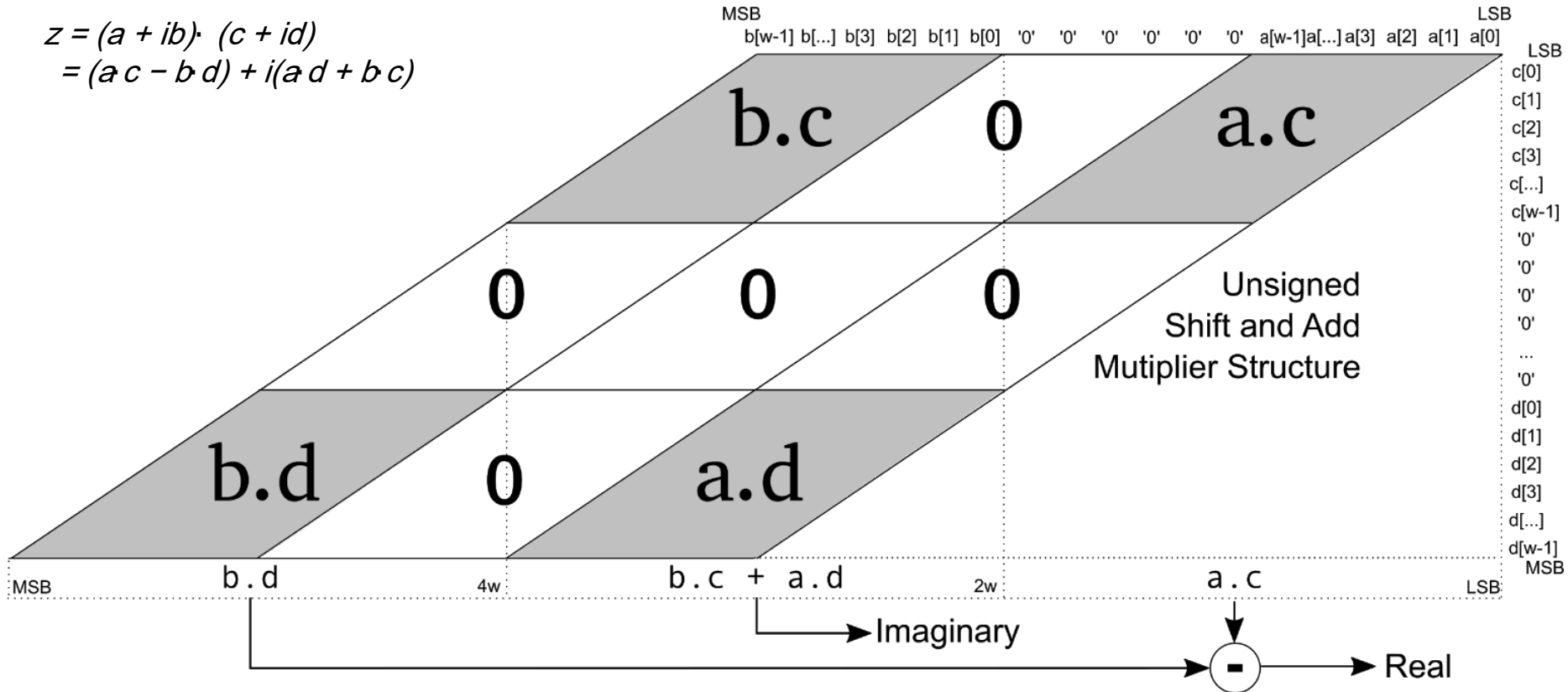
$$z = b d - a c + i(b c + a d)$$

Unsigned Complex Multiplication

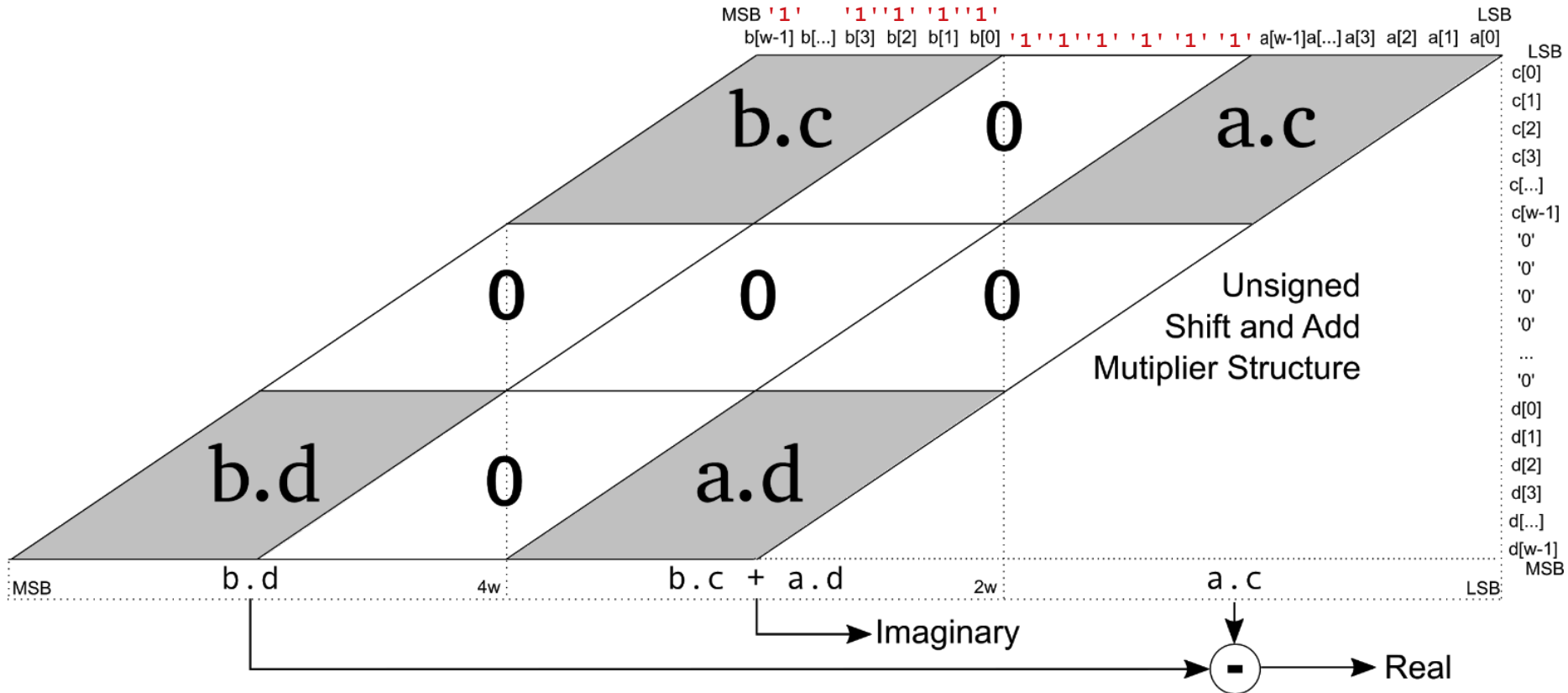


$$z = (a + ib) \cdot (c + id)$$

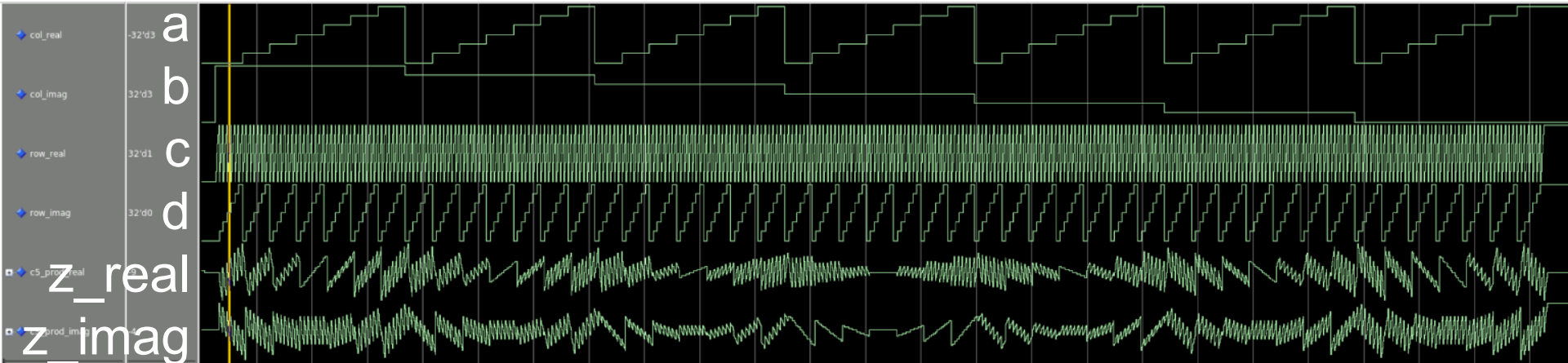
$$= (ac - bd) + i(ad + bc)$$



Signed Complex Multiplication



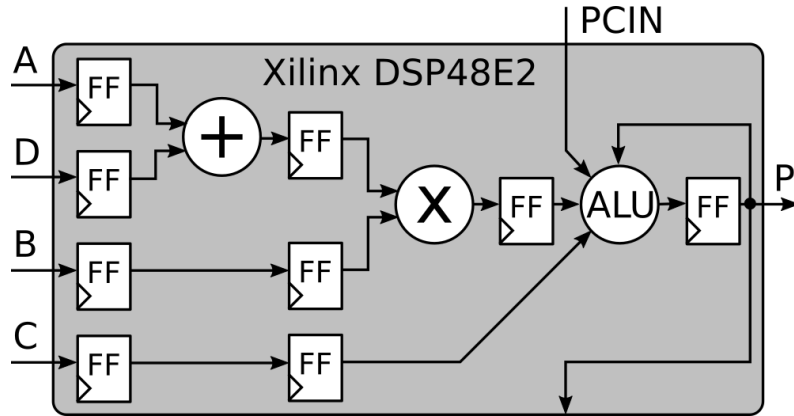
Signed Complex Multiplication



Simulation for $w=3$, therefore a, b, c, d in range -3 to $+3$.

It works! Except for the one case where $a = b = c = d = 2^w$
- but the SKA uses -2^w as NaN anyway

Ultrascale+ DSPs



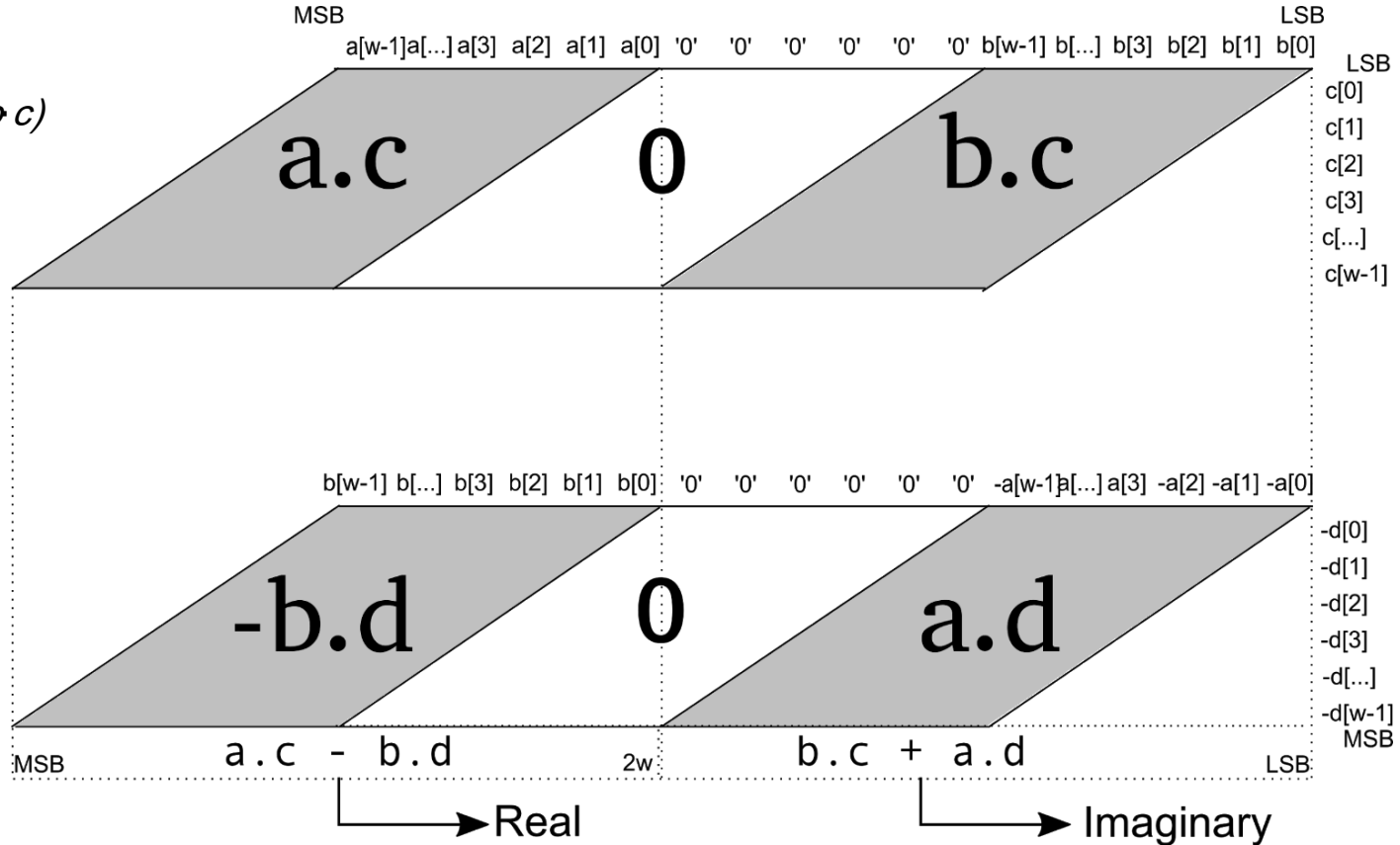
Xilinx Ultrascale+
One 18x27b signed multiply.

Ultrascale+ DSPs



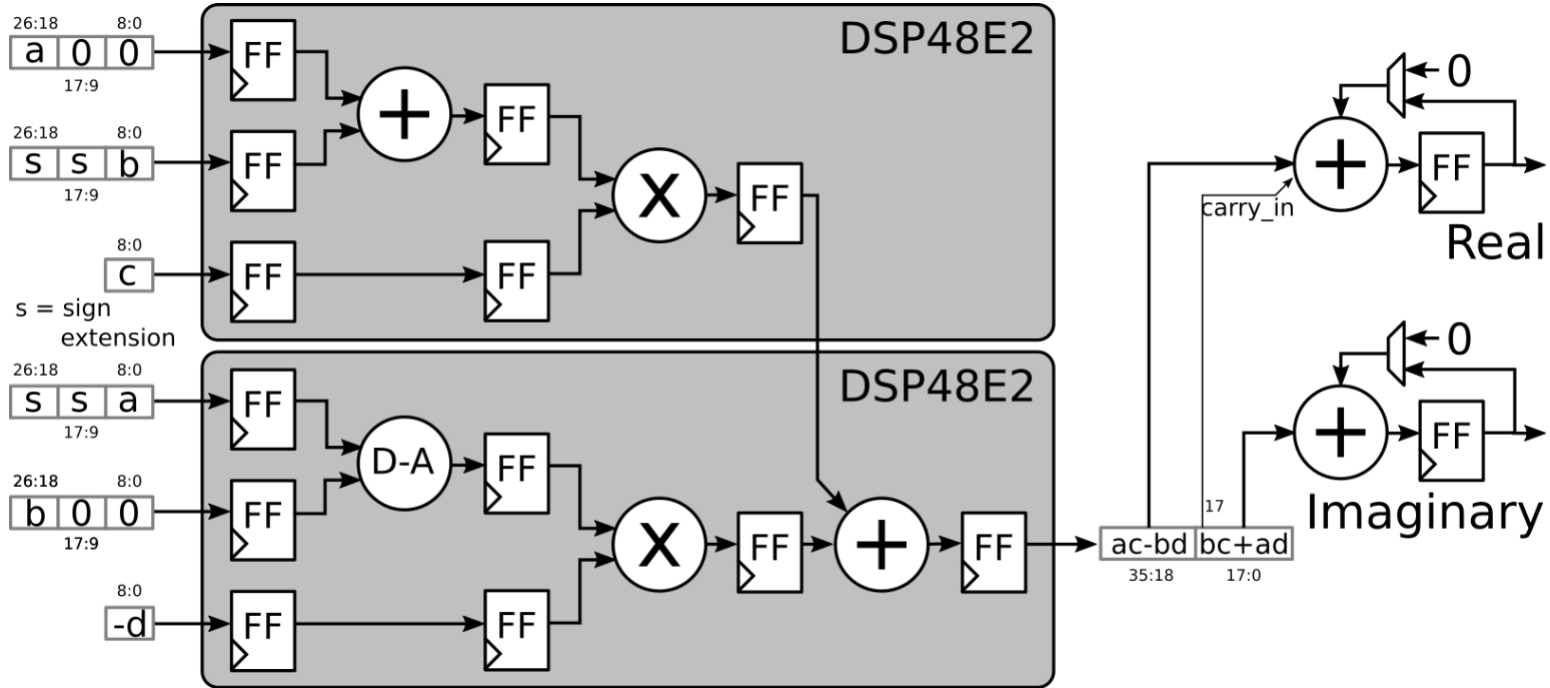
$$z = (a + ib) \cdot (c + id)$$

$$= (ac - bd) + i(ad + bc)$$



Ultrascale+ DSPs

$$z = (a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc)$$



Ultrascale+ DSPs

$$z = (a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc)$$

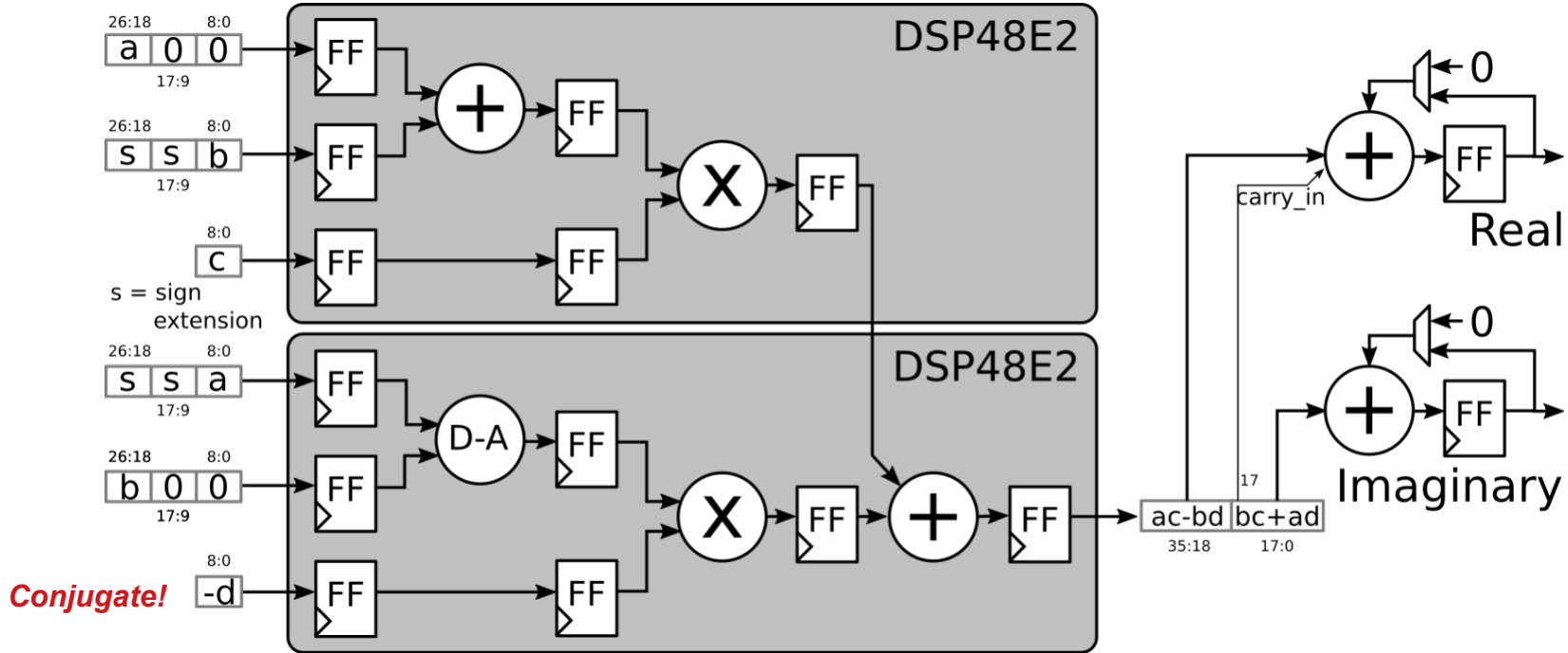


TABLE II
IMPLEMENTATION RESULTS FOR XILINX ULTRASCALE+

Style	Width	LUTs	FFs	DSPs	Fmax
optimised	9-bit	86	54	2	640 MHz
inferred	9-bit	99	163	2	450 MHz

Results: M=19, f=384 MHz



The screenshot displays the Xilinx ISE environment. On the left, the 'Sources' window shows a project named 'cmac_array' with various components like 'Nets (47355)', 'Leaf Cells (356)', and several 'E_ARRAY_DRIVER' and 'G_CMACS_ROW' components. Below it, the 'Source File Properties' window shows 'cmac_tdm_pkg.vhd' as a 'VHDL 2008' file in the 'work' library.

The main window shows a timing diagram for a grid of components labeled X000 through X715. The diagram displays signal waveforms for each component, with some components highlighted in yellow.

At the bottom, the 'Timing' window is open, showing the 'Design Timing Summary'.

Setup		Hold		Pulse Width	
Worst Negative Slack (WNS):	0.210 ns	Worst Hold Slack (WHS):	0.026 ns	Worst Pulse Width Slack (WPWS):	0.758 ns
Total Negative Slack (TNS):	0.000 ns	Total Hold Slack (THS):	0.000 ns	Total Pulse Width Negative Slack (TPWS):	0.000 ns
Number of Failing Endpoints:	0	Number of Failing Endpoints:	0	Number of Failing Endpoints:	0
Total Number of Endpoints:	943536	Total Number of Endpoints:	943536	Total Number of Endpoints:	330840

All user specified timing constraints are met.

Results: M=18, f=440 MHz



The screenshot displays a design tool interface with several panels:

- Sources:** A tree view showing the project hierarchy, including "cmac_array", "Nets (42721)", "Leaf Cells (343)", and various driver and reset pipe components.
- Source File Properties:** A panel for "cmac_tdm_pkg.vhd" showing it is enabled, located at "/home/abel/Repository/wk_bitbucket", and is of type "VHDL 2008".
- Timing Diagram:** A large grid of waveforms showing signals for various components like X100 through X174, with columns labeled SLR0 through SLR4.
- Design Timing Summary:** A table summarizing timing constraints.

Setup	Hold	Pulse Width
Worst Negative Slack (WNS): 0.106 ns	Worst Hold Slack (WHS): 0.026 ns	Worst Pulse Width Slack (WPWS): 0.568 ns
Total Negative Slack (TNS): 0.000 ns	Total Hold Slack (THS): 0.000 ns	Total Pulse Width Negative Slack (TPWS): 0.000 ns
Number of Failing Endpoints: 0	Number of Failing Endpoints: 0	Number of Failing Endpoints: 0
Total Number of Endpoints: 848818	Total Number of Endpoints: 848818	Total Number of Endpoints: 297723

All user specified timing constraints are met.

Results: M=17, f=505 MHz



The screenshot displays the Xilinx ISE software interface. On the left, the 'Sources' window shows a project named 'Netlist' with a hierarchy of files including 'cmac_array', 'Nets (38314)', 'Leaf Cells (336)', and various driver and pipeline delay files. Below this, the 'Source File Properties' window is open for 'cmac_tdm_cache.vhd', showing it is enabled, located at '/home/abel/Repository/wk_bitbucket', and is a VHDL 2008 file of size 14.8 KB.

The main workspace shows a timing diagram for a device. The diagram is a grid of signals labeled X000 through X071. The signals are grouped into columns labeled SLR0 through SLR2. The diagram shows signal transitions over time, with some signals highlighted in yellow.

At the bottom, the 'Timing' window is open, displaying the 'Design Timing Summary' table:

Setup		Hold		Pulse Width	
Worst Negative Slack (WNS):	0.064 ns	Worst Hold Slack (WHS):	0.026 ns	Worst Pulse Width Slack (WPWS):	0.448 ns
Total Negative Slack (TNS):	0.000 ns	Total Hold Slack (THS):	0.000 ns	Total Pulse Width Negative Slack (TPWS):	0.000 ns
Number of Failing Endpoints:	0	Number of Failing Endpoints:	0	Number of Failing Endpoints:	0
Total Number of Endpoints:	764686	Total Number of Endpoints:	764686	Total Number of Endpoints:	271966

Results: M=16, f=534 MHz



The screenshot displays a timing analysis tool interface. On the left, a netlist tree shows the hierarchy of components, including 'cmac_array', 'Nets (34182)', and various 'G_CMACS_ROW' and 'E_CMACS_QU' blocks. Below the netlist is a 'Properties' panel. The main window shows a timing diagram with a grid of signal waveforms labeled with component names like 'X010_X011', 'X018_X019', etc. The bottom panel shows the 'Design Timing Summary' for 'Timing'.

Setup	Hold	Pulse Width
Worst Negative Slack (WNS): 0.070 ns	Worst Hold Slack (WHS): 0.021 ns	Worst Pulse Width Slack (WPWS): 0.393 ns
Total Negative Slack (TNS): 0.000 ns	Total Hold Slack (THS): 0.000 ns	Total Pulse Width Negative Slack (TPWS): 0.000 ns
Number of Falling Endpoints: 0	Number of Falling Endpoints: 0	Number of Falling Endpoints: 0
Total Number of Endpoints: 678015	Total Number of Endpoints: 678015	Total Number of Endpoints: 237190

All user specified timing constraints are met.

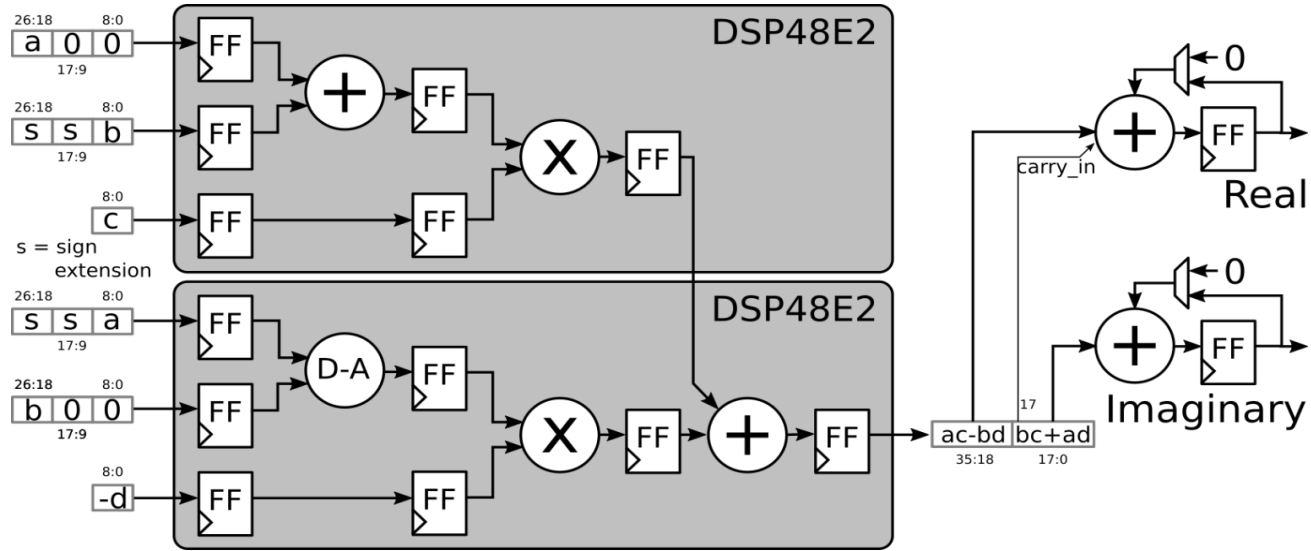
Results



Matrix Size (M)	Stations per Row (D)	TDM Slots	Stations Supported (>N/2)	TDM Cache BRAM36s	CMACs	DSP48s	Required Fmax
30 (15 Stations)	35	614	525	150	930	1860	639.6 MHz
32 (16 Stations)	32	512	512	160	1056	2112	533.4 MHz
34 (17 Stations)	31	482	527	136	1190	2380	502.1 MHz
36 (18 Stations)	29	422	522	144	1332	2664	439.6 MHz
38 (19 Stations)	27	366	513	152	1482	2964	381.25 MHz
40 (20 Stations)	26	338	520	160	1640	3280	352.1 MHz

Factor 1/2 of Karatsuba

Factor 2/3 of Karatsuba



Questions / Discussion?

Thank-you!