SKA Signal Processing Concept Design Review
Manchester, 2011 April 15

Low-Power Architectures and ASIC-Based Concept for SKA Phase 1

Larry R. D'Addario, JPL/Caltech

© 2011 California Institute of Technology

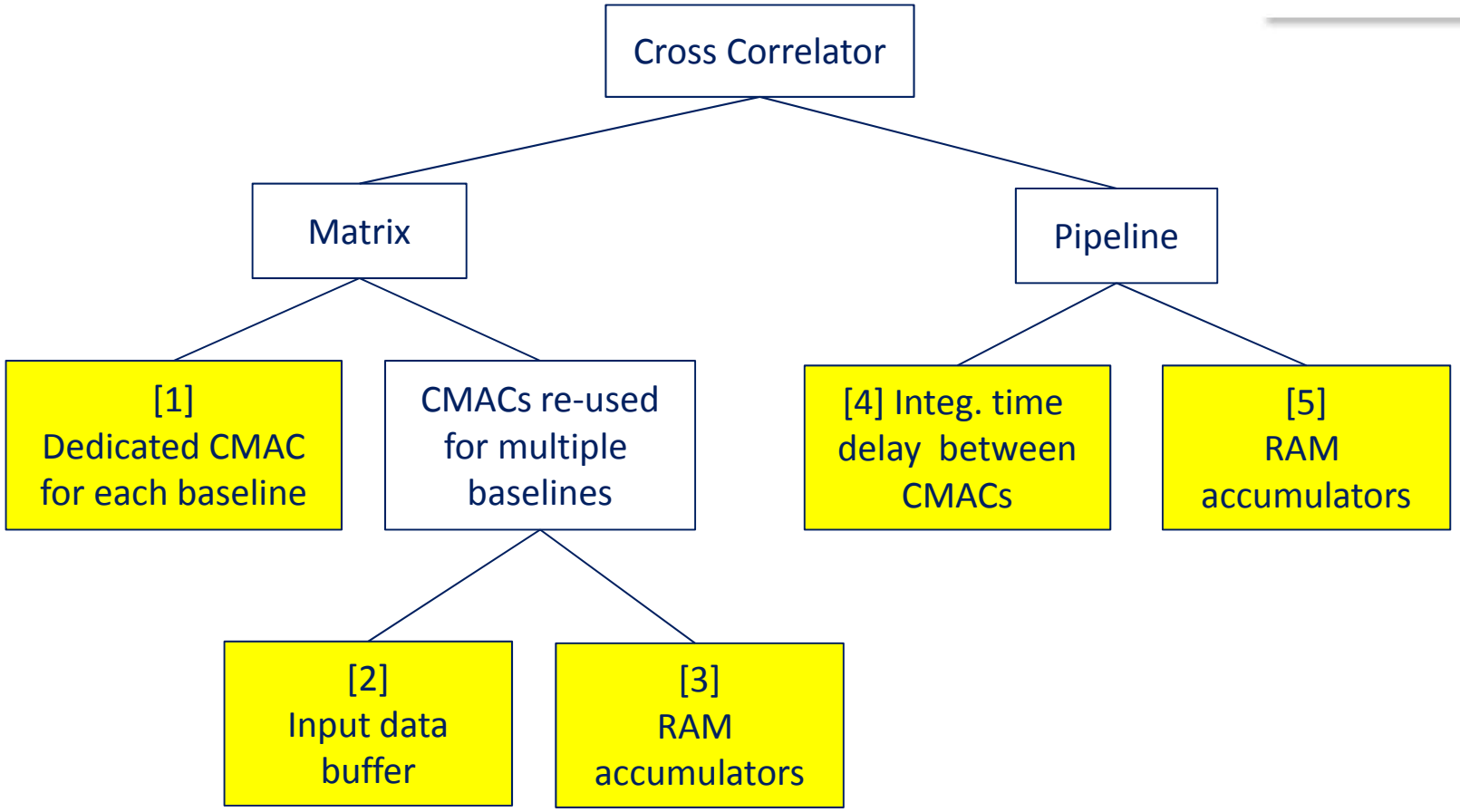
Outline

- Cross correlators in general (see SKA Memo 133 = 3f2)
  - Architectures: minimizing power
  - Use of ASICs; strawman design in 90nm CMOS.
- SKA Phase 1 (see CoDR document 3f1)
  - Both array types
    - cross correlators
    - filter banks
  - Array of dipole-array stations ("sparse aperture arrays")
    - station beamformers
    - array architecture options
- SKA Phase 2
  - Projection to more advanced technology
- Demo ASIC design at JPL

Cross Correlator Architectures

- First consider just the cross-correlator proper (ignoring all antenna-based processing), since it dominates signal processing at large *N*.
- It turns out that the size, cost, and power consumption of very-large-*N* cross-correlators are strongly dependent on architecture.
  - Rate of complex multiply-accumulate (CMAC) operations is the same in all architectures: $B\,N\,(2N+1)$ for $2N$ signals at bandwidth B.
  - But memory and I/O are also very important. Estimated energy per op in 90 nm CMOS:
    - CMAC for 2b+2b samples 2 pJ
    - Move one sample between chips 8 pJ
    - Write and read one sample to/from RAM 33 pJ
  - Memory operations dominate power and memories dominate chip area in some architectures; those architectures must be avoided.
- If the right architecture is chosen, an SKA-scale correlator can be built with current technology. There is no need to wait for Moore's Law.

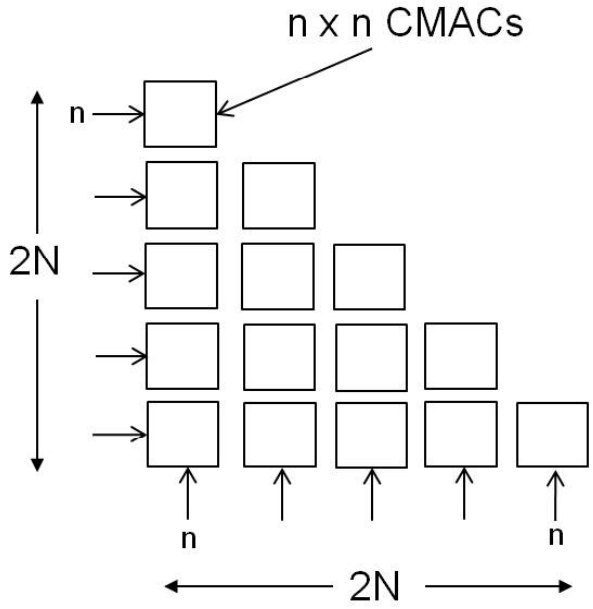
Cross Correlator Family Tree



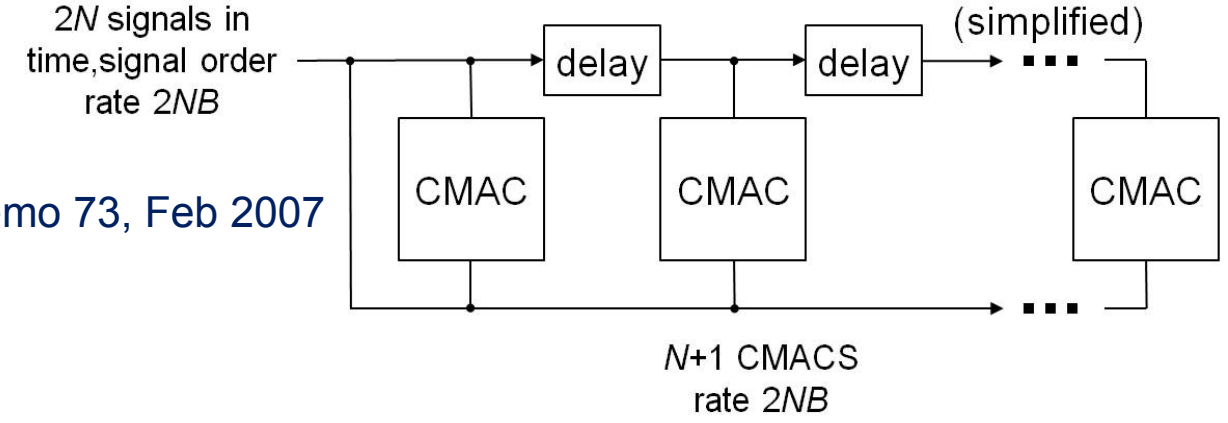
Notes:
[1] ALMA, EVLA, SKA Memo 127
[4] ATA Memo 73

Matrix vs. Pipeline

Matrix:

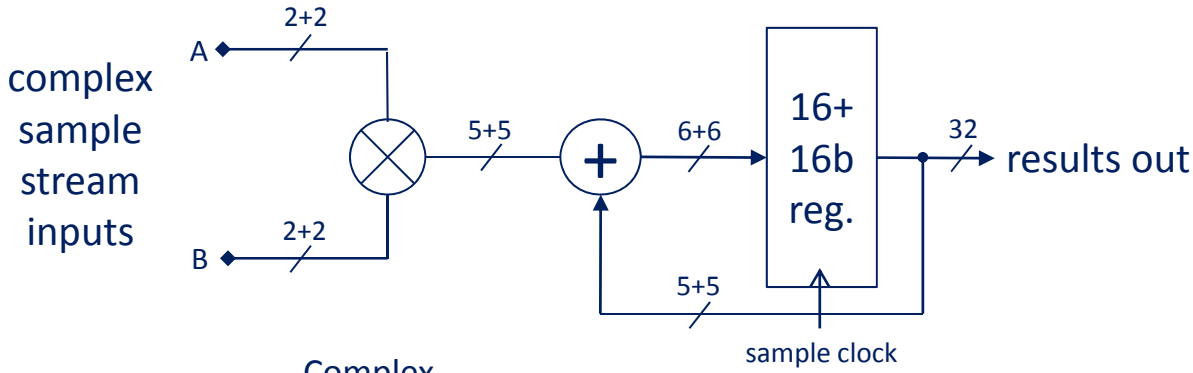


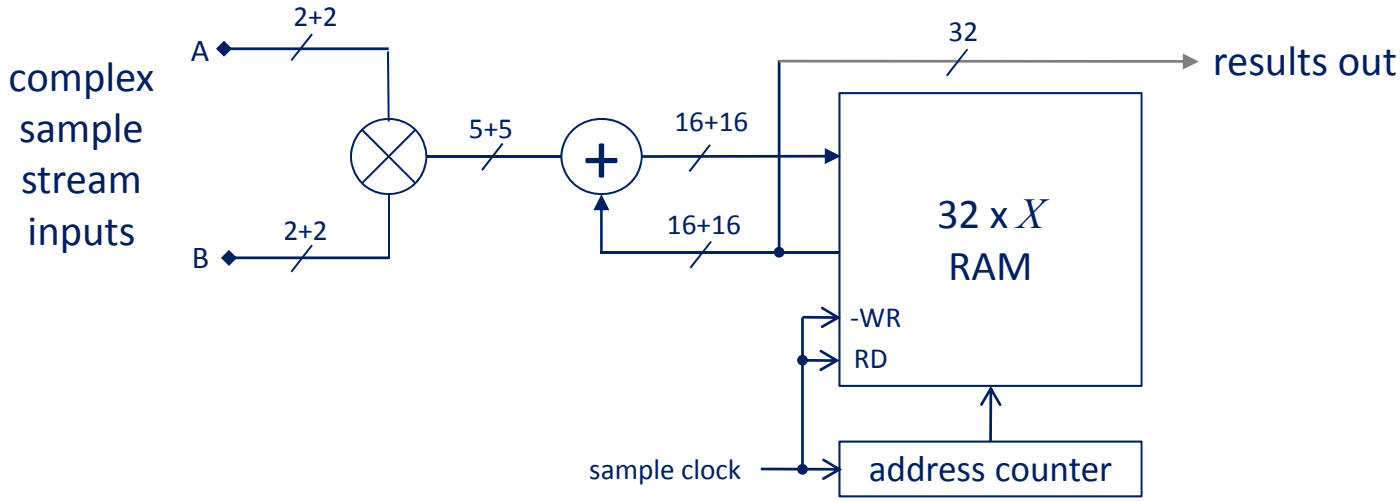
Pipeline:



L. Urry, ATA Memo 73, Feb 2007

CMAC: Dedicated vs. RAM Accumulator





Design Approach

- Find a practical implementation of an *N*≈2000, *B*=1GHz correlator for each architecture separately. This requires generating a top-level ASIC design for each case.
- Design each ASIC in 90 nm CMOS technology, subject to certain constraints.
- Find the power dissipation of the ASIC from its clock speed, the energy per operation of its main components, and the count of those components:
  - CMACs
  - Memory
  - I/O
- Determine the energies by scaling from existing designs, by simulations, and from the literature.
- Find the total power dissipation by multiplying the power per ASIC by the number of ASICs needed to implement the strawman correlator.

Chip Level Constraints

- Maximum power dissipation 75 W
- Maximum chip area 200 mm^2
- Maximum input rate 40 Gb/s
- Maximum output rate 40 Gb/s
- Comparison:
  - Nvidia's GeForce GTX 280 GPU (65 nm technology) is 576 mm^2 and dissipates up to 236 W
    http://en.wikipedia.org/wiki/Comparison_of_Nvidia_graphics_processing_units

Other parameters of straw man correlator:

- Sample quantization 2b real + 2b imaginary
- Accumulator readout 16b+16b
- Channel bandwidth 100 kHz
- Minimum integration 65 ms

Technology Parameters: Areas and Energies

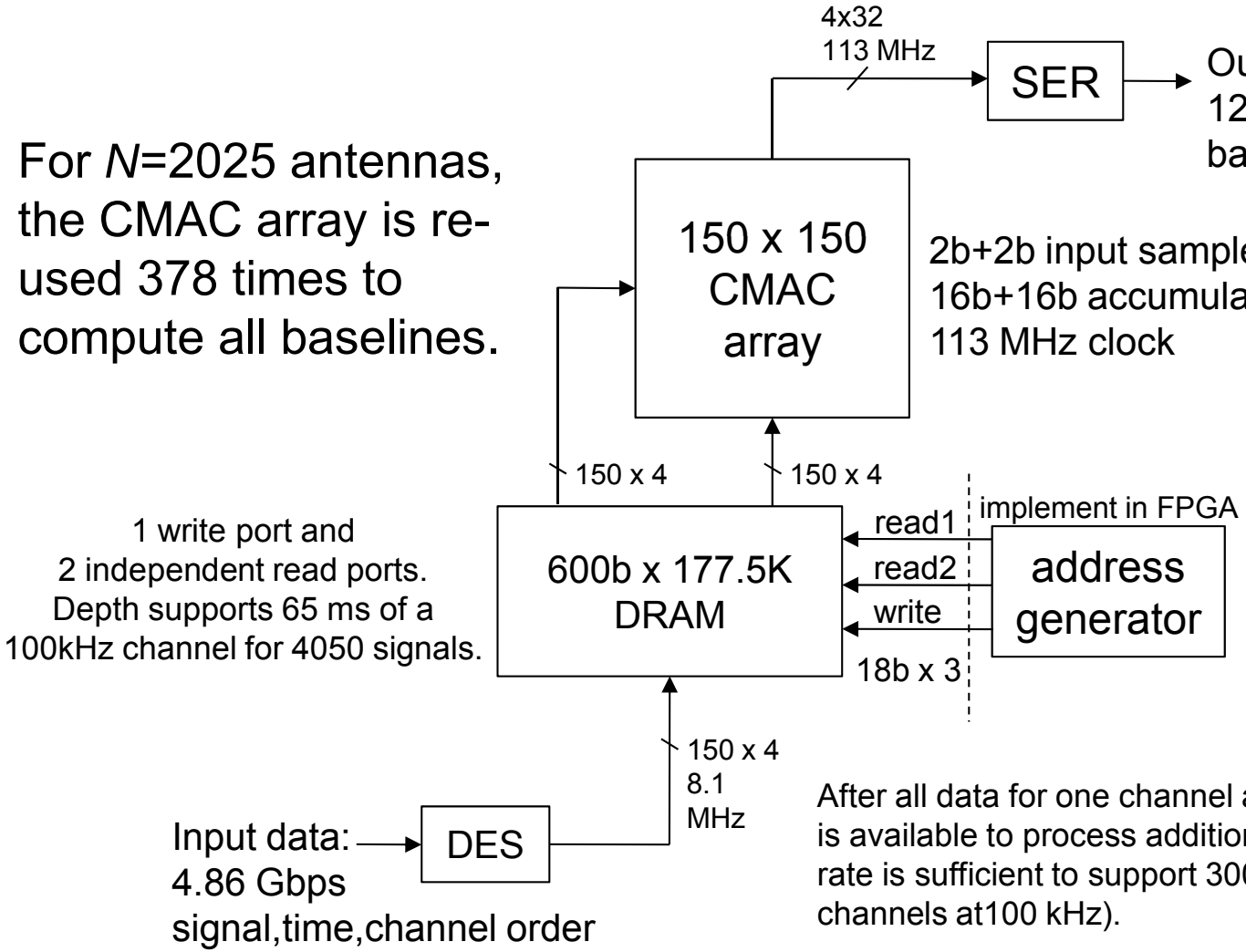
Element Model Parameters For 90 nm Process Technology				
Element (architecture)	*Die Area* mm^2	*Energy/op* pJ	*Static power* W	*Max. clock* MHz
CMAC (all)	0.006303	2.45	0	N/A
Transceiver, 6.25 Gb/s (all)	0.266	2.12	0	N/A
DRAM 600x177.5k (#2)	48.7	3750	0.593	127
SRAM 32x100 (#3)	.0130	1.87	.000143	1165
SRAM 4x6500 (#4)	.0488	2.97	.000102	695
SRAM 32x4000 (#5)	0.2428	14.0	.00499	661

Summary of Results (90 nm Technology)

Architecture:	*1*	*2*	*3*	*4*	*5*
Number of antennas, N	2024	2025	2040	2000	2000
CMACs/IC, m	30976	22500	10404	2001	667
Channels/IC, K	284	3	24	1	1
Clock frequency, f, Hz	2.84E+07	1.13E+08	2.40E+08	4.00E+08	4.00E+08
Input rate/IC, b/s	4.00E+10	4.86E+09	3.92E+10	6.40E+09	6.40E+09
Output rate/IC, b/s	4.33E+09	1.26E+10	1.23E+10	1.03E+10	7.71E+09
Power/IC, W	2.25	7.73	17.05	11.49	11.50
Die area, mm^2	195.24	190.50	200.51	208.06	166.18
Memory power, %	0.0%	18.7%	63.5%	82.6%	94.1%
I/O power, %	4.2%	0.5%	0.6%	0.3%	0.3%
Memory area %	0.0%	25.6%	67.3%	93.9%	97.5%
ICs to process all baselines, c_1	276	1	8	1	3
Total ICs in system, c	9,718	3,333	3,333	10,000	30,000
Total power, all ICs, W	21,859	25,771	56,823	114,873	344,873

ASIC Design for Architecture #2

For *N*=2025 antennas, the CMAC array is re-used 378 times to compute all baselines.



After all data for one channel are processed, time is available to process additional channels . Clock rate is sufficient to support 300 kHz bandwidth (3 channels at100 kHz).

SKA Phase 1 Concept

- Use the ASIC just described to build the correlators
  - Merely conceptual, not necessarily what we would really do
  - A smaller ASIC would be more cost effective at the smaller *N* of SKA1
- Adjust the SKA Memo 130 parameters for better efficiency
  - Match to ASIC
  - For SKA1-low, consider beam formers and minimize overall power

Correlator Requirements, Memo 130

(adjusted values)	*Reflector Antennas*	*Dipole-like Elements*
N, number of antennas or stations	250 (225, 300)	50 (75, 150)
B, total bandwidth, MHz	1000	380
b, channel bandwidth, MHz	.015	.001 to 0.05 (.01)
J, number of station beams	1	480 (320, 160)
τ, integrating time, minimum, s	0.1	1.2
τ, integrating time, maximum, s	4.1	49

Correlators for SKA1

	Dishes			*Dipoles*		
	N =300	*N* =300	*N* =225	*N* =150	*N* =150	*N* =75
	min τ	max τ	max τ	min τ	max τ	max τ
Parameters						
Number of antennas or stations	300	300	225	150	150	75
Number of station beams	1	1	1	160	160	320
Total bandwidth, Hz	1.00E+09	1.00E+09	1.00E+09	3.80E+08	3.80E+08	3.80E+08
Channel bandwidth, Hz (average)	1.50E+04	1.50E+04	1.50E+04	10000	10000	10000
Integration time, short term, s	0.1	2.9	3.9	1.2	8.5	17
Chip level results						
Channels/chip	550	550	920	3300	3300	6600
Re-use factor	10	10	6	3	3	1
Clock rate, Hz	8.25E+07	8.25E+07	8.28E+07	9.90E+07	9.90E+07	6.60E+07
Input rate, b/s	1.98E+10	1.98E+10	2.48E+10	3.96E+10	3.96E+10	3.96E+10
Output rate, b/s	3.96E+10	1.37E+09	1.02E+09	5.94E+09	8.39E+08	2.80E+08
Chip power, total, W	5.89	5.80	5.83	6.89	6.88	4.81
System level results						
Chip count, total in system	121	121	72	1,842	1,842	1,842
Total power, all chips, W	713	704	423	12,693	12,673	8,863

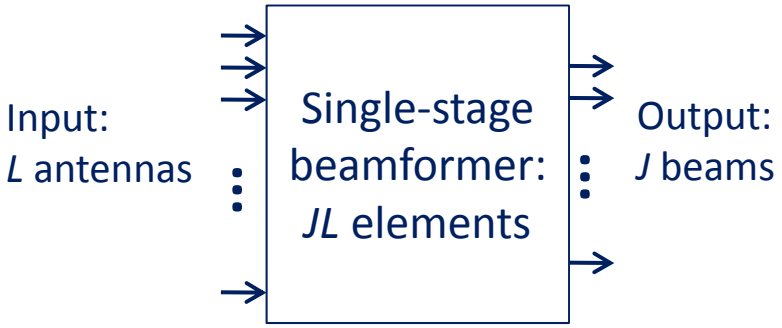
Filter Banks

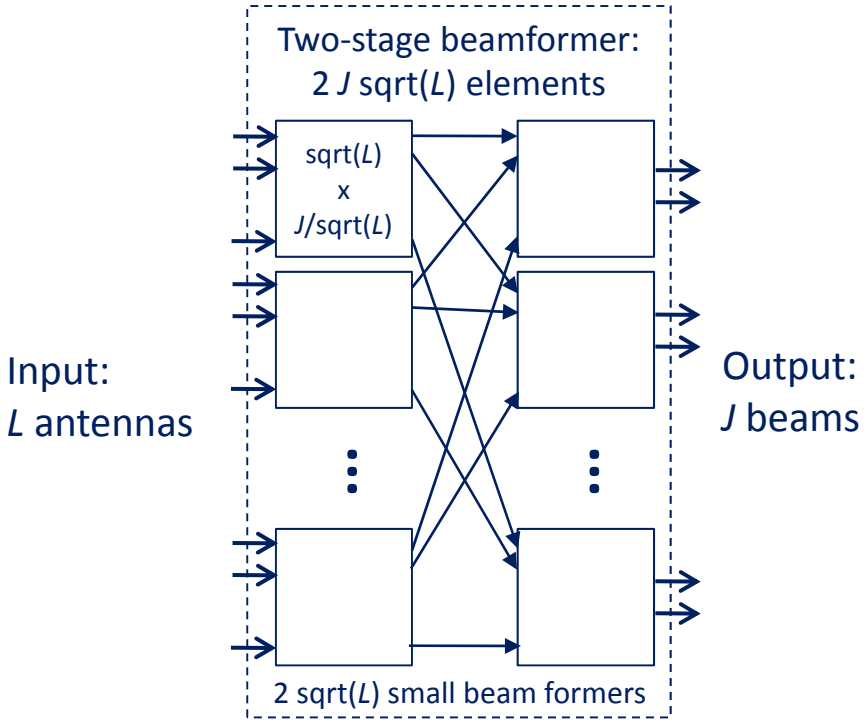
- Richards et al., *IEEE Custom Integrated Circuits Conference*, Sep 2009:
  - ASIC implementing a 4096 channel spectrometer for 750 MHz bandwidth
  - Built in 90 nm CMOS, 7.8 mm^2 die, dissipates 0.71 W.
  - 9.75 $x10^9$ FFT butterflies/second
  - 62.6 pJ per butterfly operation
  - We could put 8 of these on a 62 mm^2 die → 7.80 $x10^{10}$ butterflies/chip
- Each filter bank requires $(B/2)\log_2(B/b)$ FFT butterflies/second.
  - Ignore polyphase pre-filters when B/b is large.
  - Dishes: B/b = 66,667; round to 65,536
  - Dipole arrays: B/b = 38,000 (channels 4 kHz to 23 kHz, 10 kHz average).
- Number of filter banks needed
  - Dishes: two per antenna per beam (=2N for SKA1)
  - Dipole arrays: two per station element or two per beam

    = $2Ne_s$ or $2Nb$

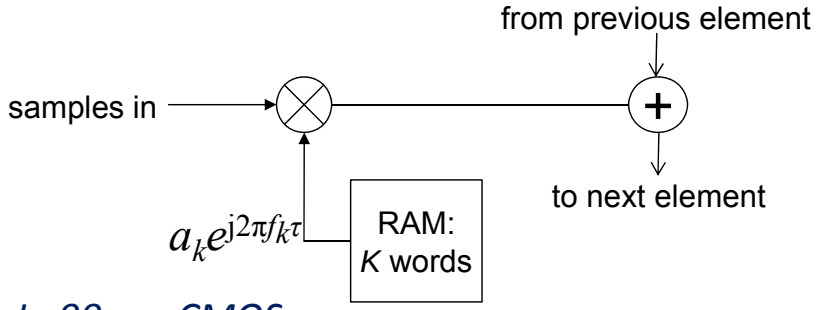
    depending on whether filter banks go before or after beam formers.

Station Beam Formers





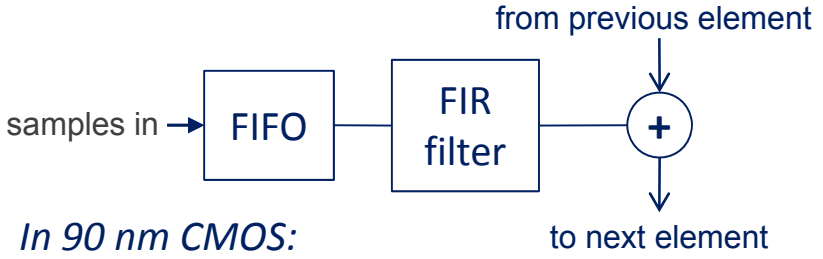
Frequency Domain Beam Forming Element



In 90 nm CMOS:

- 12 pJ/operation
- 256 elements in a 200 mm^2 ASIC at K=65536

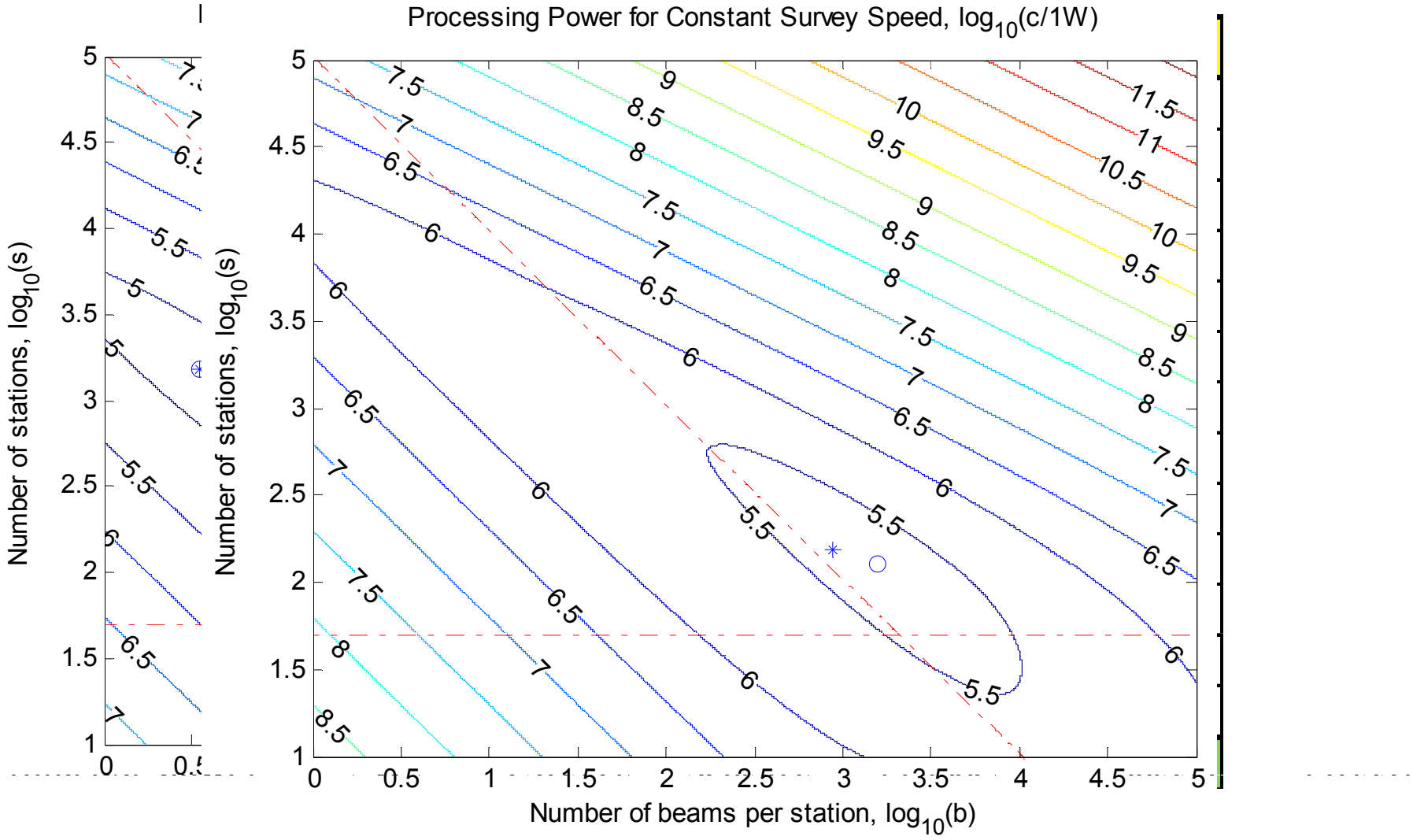
Time Domain Beam Forming Element



In 90 nm CMOS:

- 80 pJ/operation
- 256? elements in a 200 mm^2 ASIC

SKA1-low Options



Exploring the Universe with the world's largest radio telescope

SKA2 Correlator ASIC

	SKA2-mid	
	90 nm	32 nm
Number of CMACs	25600	102,400
Memory size, Mbits	80	320
Clock frequency, MHz	125	750
Maximum number of antennas	2560	10240
Bandwidth @ Nmax, MHz	0.2367	0.3606
Die area, mm^2	208.4	127.4
Actual N	2025	2080
Bandwidth, MHz	0.3754	8.2418
Integrating time, samples	5178	20165
Input rate, Mbps	6,081	137,143
Output rate, Mbps	19,775	121,875
Chip count for B=1GHz	**2665**	**122**
Total power, W	**21,908**	**8,610**

-- backup slides follow --

CMAC Energy Scaling, 3 cases

- ALMA (~2001)
  - 250 nm, 1.8V, 4096 real MACs, 1.98W, 125 MHz => 3.9 pJ /MAC op.
  - 2b samples, 16b accumulators (same as our strawman).
  - No RAM; I/O power is negligible.
  - Assume CMAC/MAC = 3.5 in energy 13.5 pJ/CMAC op.
  - Current scales as gate length, Vdd for 90nm is 0.9V
    x(90nm/250nm)(0.9V/1.8V) 2.45 pJ/CMAC op.
- EVLA (~2004)
  - 130 nm, 1.02V, 2048 "complex" MACs, 1.88W core, 256 MHz.
  - 4b real samples in, VLBI-style architecture with fringe rotation.
  - Assume net scaling by 2/3 for complexity from 4b VLBI CMAC to 2b+2b true CMAC.
  - No RAM; I/O power separately accounted (0.7W/chip).
  - Technology scaling: x(90nm/130nm)(0.9V/1.08V)
  - Net result 1.46 pJ/CMAC op
- GeoSTAR spacecraft (prototype chip, 2009)
  - 90 nm, 289 CMACs, 2b+2b, 1 GHz, 1mW/cell 1.0 pJ/CMAC op

Comparison of SKA1 Concepts

Rough attempt to compare power and cost of designs in concept papers. Should not be taken too seriously, since designs did not all stick to the Memo 130 specs, and they varied considerably in their projections of future technology.

SKA Phase 1 Signal Processing Design Concepts $/euro: 1.3811

Name	Author	Correlation Only						AA Station Beamformers			Notes
		Dish Array			Sparse AA						
		chips[1]	cost[2]	power	chips[1]	cost[2]	power	chips[1]	cost[2]	power	
			k$	W		k$	W		k$	W	
ASIC based	D'Addario	121		704	1,842		12,693	22,913		178,317	300 dishes, 150 AA stns [3]
Systolic ASICs	Carlson	504	61	2,485	3,360	324	16,567	not considered			Processing boards only [4]
Uniboard based	Szomoru	2,048	3,536	89,600	24,320	41,985	1,064,000	166,400	287,269	7,280,000	AA: 16 subel/el; BF includes FBs
ASKAP-like	Bunton	100	180	2,280	640	2,600	16,000	not considered			For 2017, ~Virtex 9?
CASPER based	Kapp	2,584	16,021	262,000	7,680	47,616	3,000,000	not considered			For 2014, 'Roach 4', 'Virtex 8'.
GPU based ("sw")	kim	250	1,726	250,000	2,000	13,811	2,000,000	not considered			Based on computation only
GPU based ("sw")	Ford	752	2,596	188,000	2,256	7,789	564,000	not considered			2011 devices

[1] For the GPU based concepts, the "chips" columns contain the count of computing nodes.

[2] Costs do not include NRE.

[3] ASIC-based: 90 nm technology. Sparse AA correlator is for 150 stations, so it has 9x the capability of the others (which are each for 50 stations).

[4] Systolic ASICs: 30 nm technology. Dishes=7 boards of 72 ASICs, AAs=7chips/beam=3360chips; 14chips+aux/board->240boards.