Sieving Hardware for the NFS: Architectures and their Bottlenecks

Willi Geiselmann

The Number Field Sieve

- Precomputation
- Relation collection
- Linear Algebra (Matrix step)
- Postprocessing

Relation Collection

- Given $F_1(x,y)$, $F_2(x,y) \in Z[x,y]$, *homogeneous polynomials, e.g. of degree 5 and 1*
- Find $(a,b) \in Z \times N$ with $F_1(a,b)$ and $F_2(a,b)$ smooth, gcd(a,b) = 1

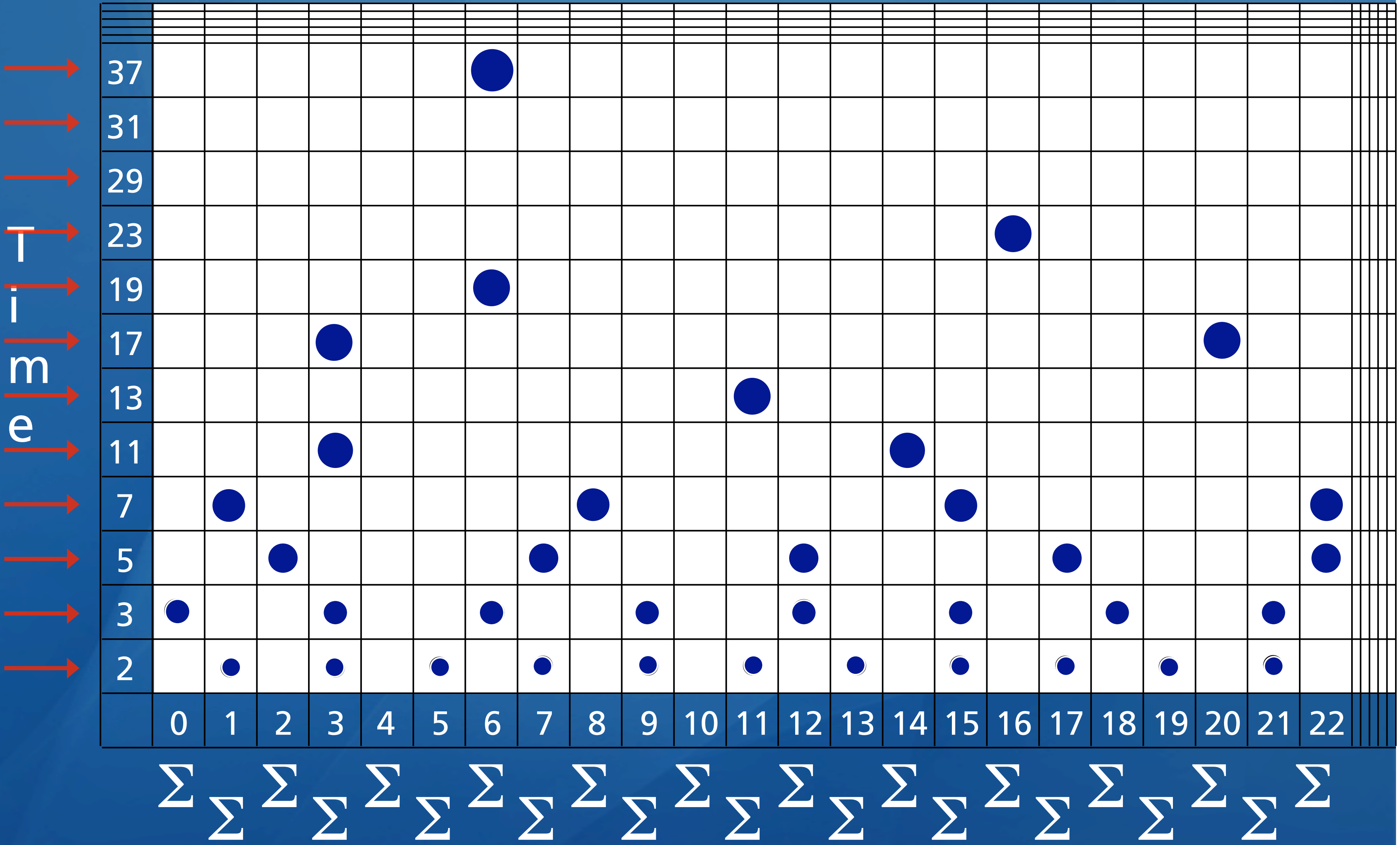
Parameters for 1024 Bit
(from TWIRL, 2003)

- Smoothness bounds:
  $B_1 = 2.6 \bullet 10^{10}$ (algebraic),
  $B_2 = 3.5 \bullet 10^9$ (rational). (≈ 2 x 5 GByte)
- Sieving region:
  $A = 5.5 \bullet 10^{14}$, $-A < a < A$;
  $B = 2.7 \bullet 10^8$, $0 < b < B$. (≈ 1400 TByte)

Suggested Hardware Designs

- TWINKLE [Shamir 1999; Shamir, Lenstra 2000]
  not designed for 1024 bit numbers
- TWIRL [Shamir, Tromer 2003]
  full wafer design
- Mesh-based sieving [G., St. 2003, 2004]
  not feasible for 1024 bit numbers
- SHARK [Franke et al. 2005]
  elaborated butterfly transport system
- SmallChips (Non-Wafer-Scale Sieving HW) [G., St. 2007]
  more realistic, but high inter-chip communication

Sieving (Eratosthenes)



Time

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
37							●																
31																							
29																							
23																	●						
19							●																
17				●																	●		
13												●											
11				●											●								
7		●							●							●							●
5			●					●					●					●					●
3	●			●			●			●			●			●			●			●	
2		●		●		●		●		●		●		●		●		●		●		●	

Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ

Memory

Sieving (TWINKLE/TWIRL)

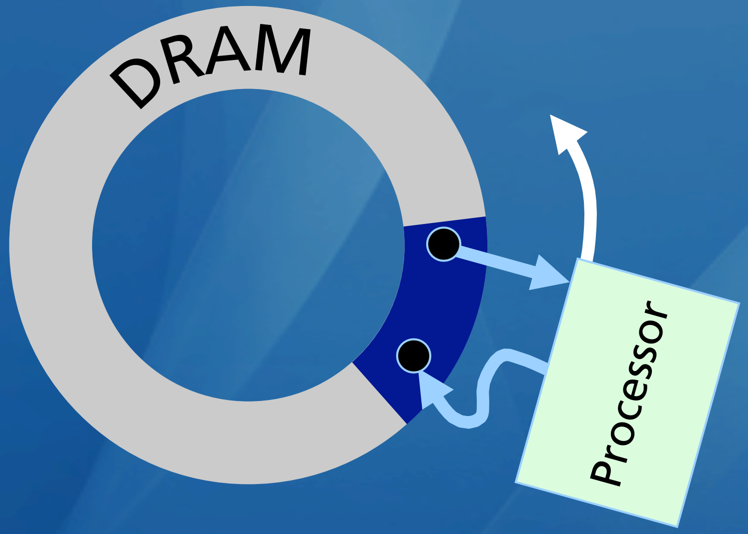
Counter	37							●																
Counter	31																							
Counter	29																							
Counter	23																	●						
Counter	19							●																
Counter	17				●																	●		
Counter	13												●											
Counter	11				●											●								
Counter	7		●							●							●							●
Counter	5			●					●					●					●					●
Counter	3	●			●			●			●			●			●			●			●	
Counter	2		●		●		●		●		●		●		●		●		●		●		●	
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ

Time

TWIRL - Types of Primes

- Largish primes (rational): $2^{19} < p$
  (algebraic): $2^{22} < p$
- Medium primes: $256 < p$
- Smallish primes: $p < 256$

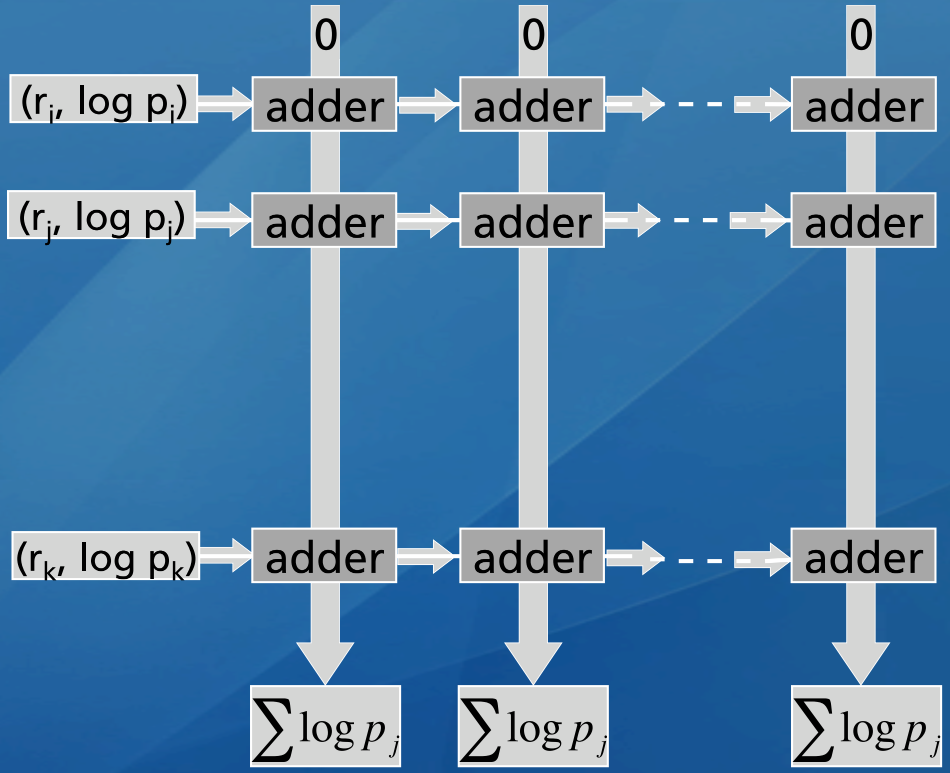
TWIRL - Largish Stations DRAM



- Rotates / reads with constant speed
- Writes to the “correct” address

TWIRL - Largish Stations
Adder Array

- Rational:
  2 100 lines
  4 096 columns
  ≈ $2^{23}/2$ adders
- Algebraic:
  14 900 lines
  32 768 columns
  ≈ $2^{29}/64$ adders



TWIRL - Med./Small Stations

- For $p < 2^{19}$ (2^{22})
- Stations are different:
  smaller DRAM,
  more logic to generate “multiple hits”
- Adder Array similar (≈ 500 lines)

TWIRL - Parts/Communication (rational)

- Size: 160 cm^2

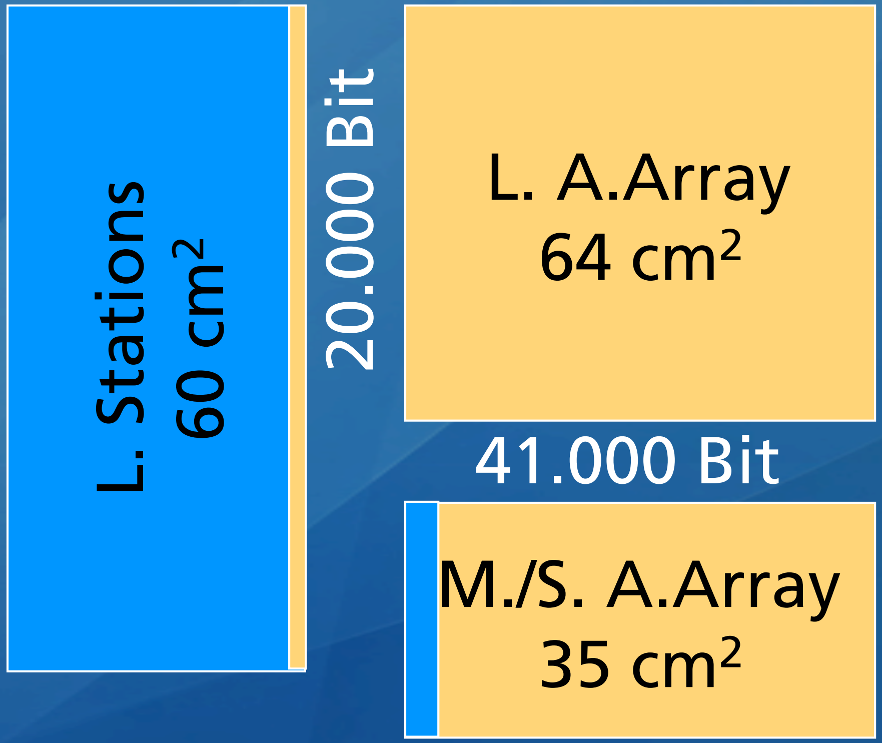
  (60 cm^2 DRAM)

- L. Stations:

  60 cm^2 (incl. DRAM)

- Adder Array:

  64 cm^2 + 35 cm^2



(when using a 0.13 μm process)

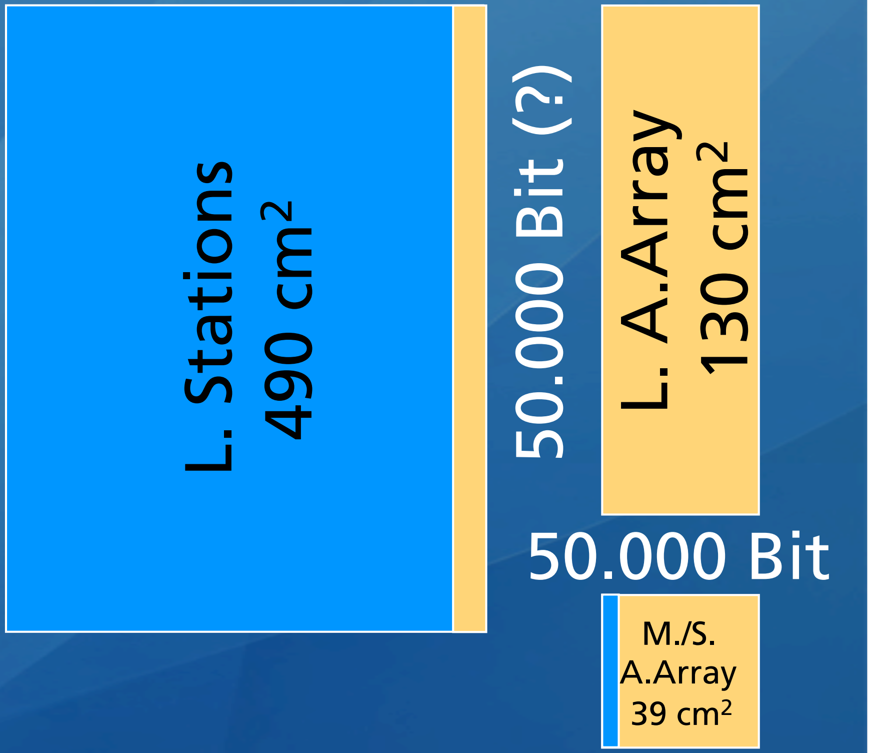
TWIRL - Parts/Communication (algebraic)

- Size: 659 cm^2

  (435 cm^2 DRAM)
- L. Stations:

  490 cm^2 (incl. DRAM)
- Adder Array:

  130 cm^2 + 39 cm^2



(when using a 0.13 μm process)

TWIRL - Performance

- Total chip area 8 x 160 cm^2 + 659 cm^2
- One sieving line in 33 sec (1 GHz)
- Sieving of a 1024 bit number with 194 devices in one year
- Fastest design (time x area)

TWIRL – Problems/Solutions

- Devices can not (easily) be cut into pieces (I/O bandwidth of chips).
- Larger Factorbasis increases this problem.
- Production errors especially in Adder Array cause problems:
  Redundancy required (increases size).
- Smaller production process and/or significant increase in I/O bandwidth would help.

Sieving (mesh / SmallChips)

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
?	37							●																
?	31																							
?	29																							
?	23																	●						
?	19							●																
?	17				●																	●		
?	13												●											
?	11				●											●								
?	7		●							●							●							●
Counter	5			●					●					●					●					○
Counter	3	●			●			●			●			●			●			●			○	
Counter	2		●		●		●		●		●		●		●		●		●		○		○	

Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ Σ

$(S = 2^{26})$

Memory

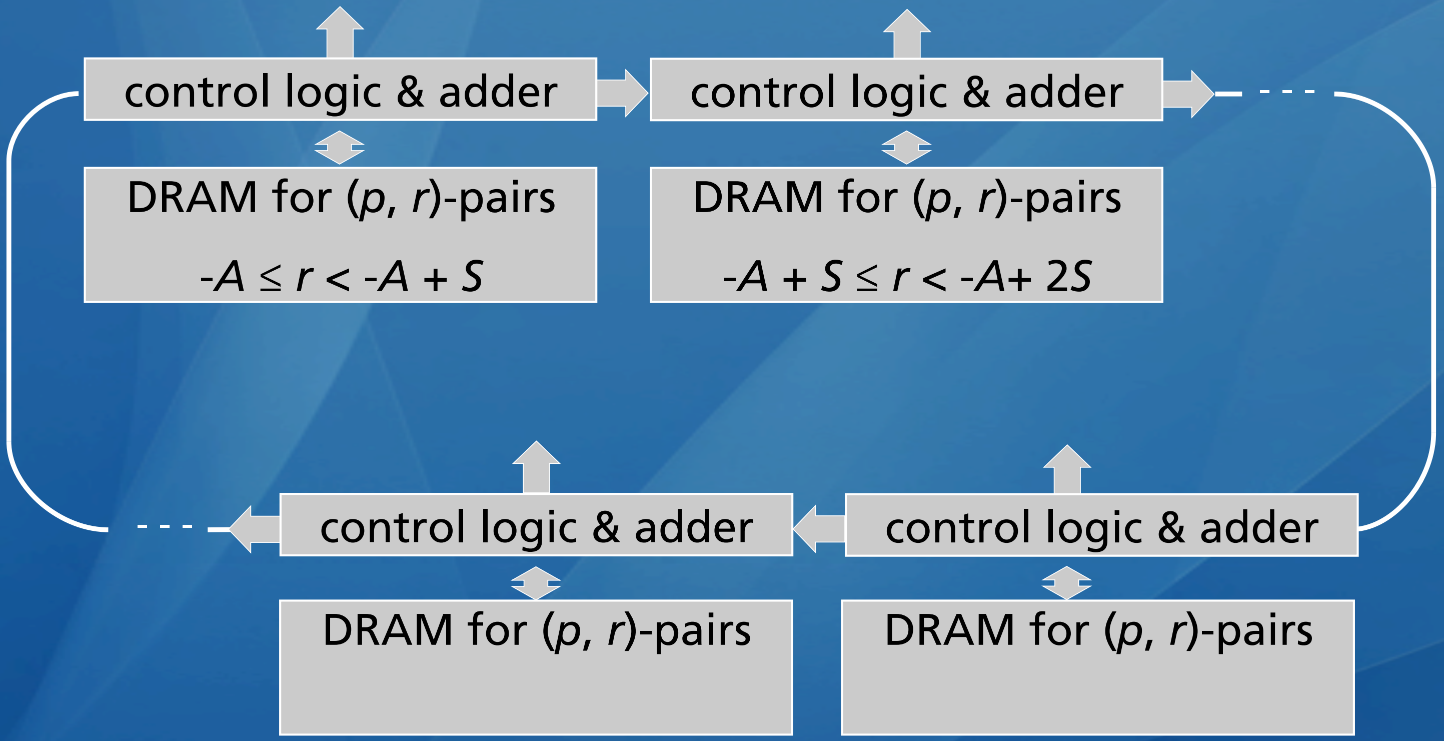
SmallChips - Idea

- Sieve intervals of a given size (2^{26})
- Generate (the rare) hits of large primes on different chips
- Collect the hits in the memory cell responsible for this sieve location

SmallChips - Types of Primes

- Largish primes I: $2^{27.2} < p < B_1 < 2^{35}$
  ...Type II/III: $1.5 \bullet 10^7 < p < 2^{27.2}$
- Medium primes: $2^{13} < p < 1.5 \bullet 10^7$
- Smallish primes: $p < 2^{13}$

SmallChips - Largish Stations



control logic & adder

DRAM for (p, r)-pairs

$-A \le r < -A + S$

control logic & adder

DRAM for (p, r)-pairs

$-A + S \le r < -A + 2S$

control logic & adder

DRAM for (p, r)-pairs

control logic & adder

DRAM for (p, r)-pairs

SmallChips - Largish Stations

- 256 stations for $p > 1.5 \bullet 10^7 \approx 2^{27.2}$
- Distributed on 32 chips:
  size: 472 mm^2 (0.13 μm process)
  output: 448 bit per clock cycle
  memory: 99%, logic: 1%
- DRAM to store both FBs: 160 cm^2

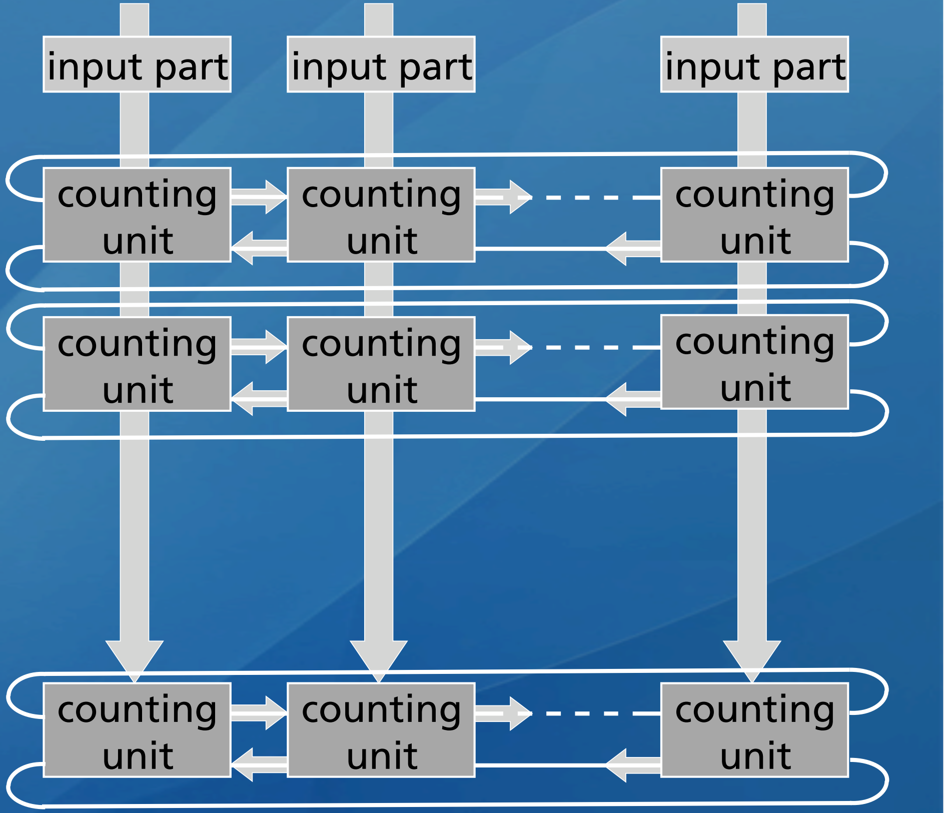
SmallChips - Medium/ Smallish Stations

Different type of storage:

- First (*p*,*r*)-pair are stored, others are calculated
- For $p < 2^{20}$: calculated in the collection unit (reduces communication, increases storage/area

SmallChips - Collection Unit

- Distributed on 4 chips, each holding
- 4 arrays of 32 x 32 counting units.
- Each unit is in charge of 2^{12} sieve locations,
- and adding up the log(p) values.



SmallChips - Collection Unit (area estimates)

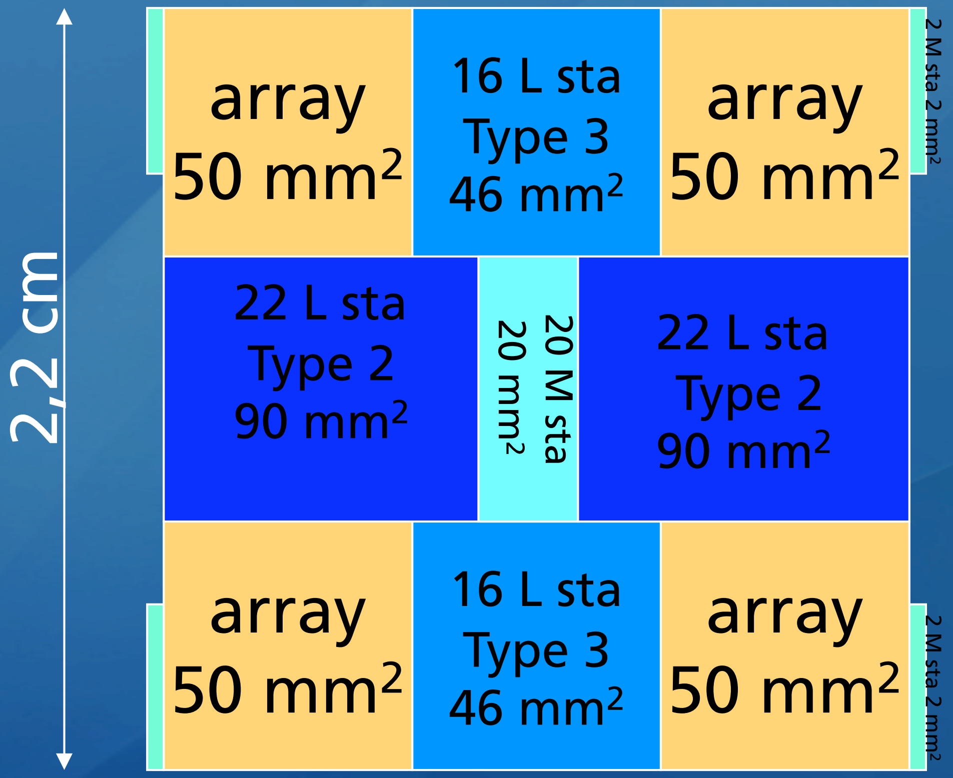
Distributed on
4 chips:

size: 493 mm²
(0.13 µm process)

input: 3584 bit / cc

memory: 94%

logic: 6%



2,2 cm

array
50 mm²

16 L sta
Type 3
46 mm²

array
50 mm²

2 M sta 2 mm²

22 L sta
Type 2
90 mm²

20 M sta
20 mm²

22 L sta
Type 2
90 mm²

array
50 mm²

16 L sta
Type 3
46 mm²

array
50 mm²

2 M sta 2 mm²

SmallChips - Performance

- Total silicon area 172 cm^2
- One subinterval ($S=2^{26}$) in 53,000 cc
- One sieve line in 25 min (600 MHz)
- Sieving of a 1024 bit number with 8300 devices in one year
- 3.5 x more silicon area than TWIRL
- or 2.0 x more after modification

SmallChips - Problems/ Solutions

- Very fast communication as input to collection unit ->
  distribute collection unit on more chips.
- Smaller process reduces chip size and/or allows to increase FB,
  communication will not increase much.
  4% FB ↔ 0.4% communication
  100% FB ↔ 10% communication

Conclusion

- SmallChips seems to be feasible
- Design/production costs are high
- Running costs are very high:
  8300 devices require 1.6 MWatt
  (200 W per device seems optimistic)
- -> 1 400 000 € per Factorization