EMICRO/SIM 2013
XV Escola de Microeletrônica Sul / 28º Simpósio Sul de Microeletrônica
Porto Alegre, 29 de abril a 3 de maio de 2013

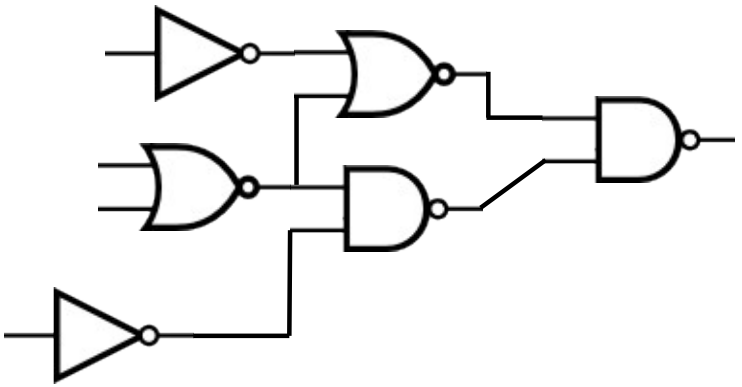
Simultaneous Gate Sizing and Vth Assignment using Lagrangian Relaxation and Delay Sensitivities

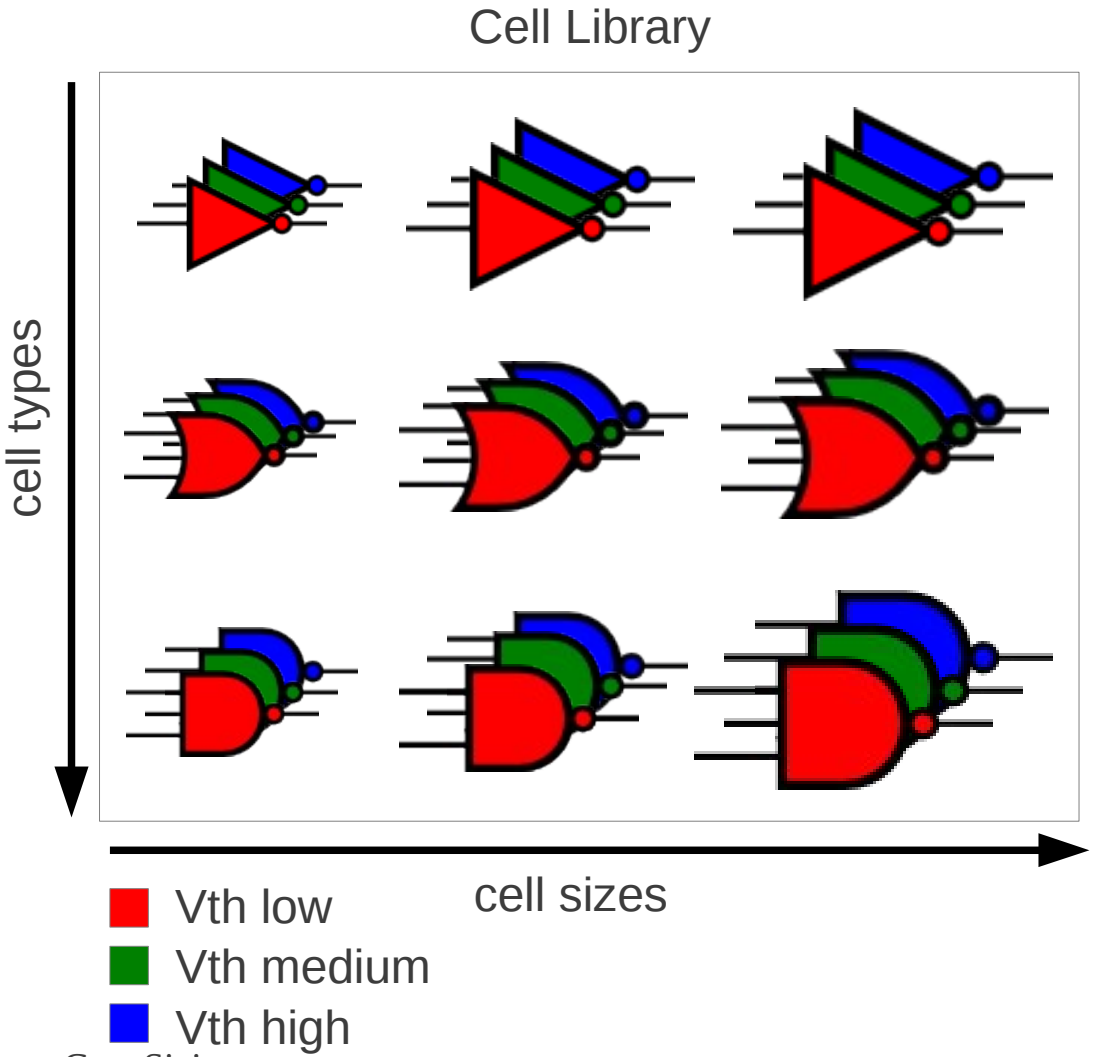
Guilherme Flach, Tiago Reimann, Gracieli Posser
Marcelo Johann, Ricardo Reis

UFRGS
PGMICRO
PROGRAMA DE
PÓS-GRADUAÇÃO EM MICROELETRÔNICA

What is gate sizing?

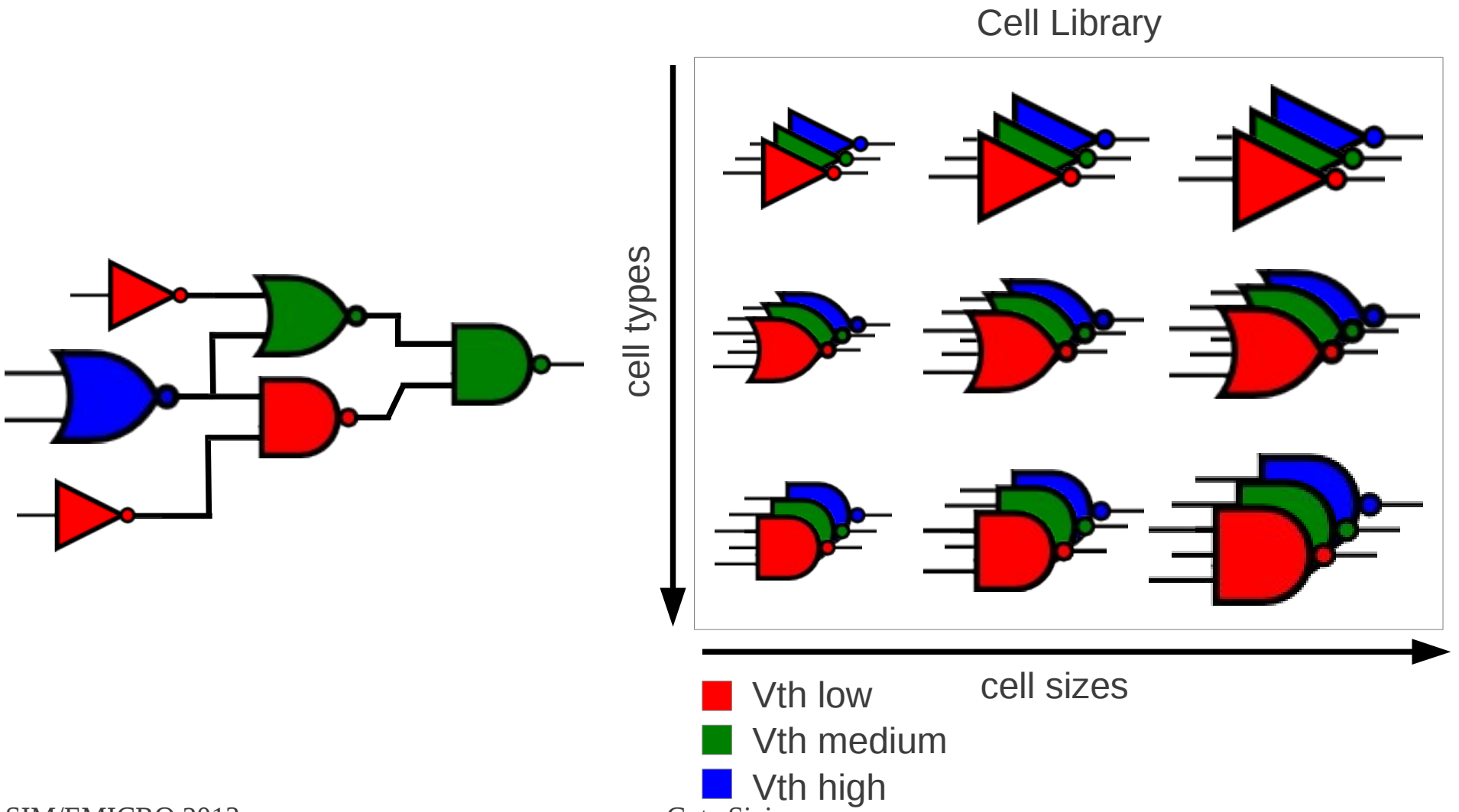
Commonly a digital circuit is assembled using standard cells from a library.





What is gate sizing?

Circuit description is then mapped to this library.



What is gate sizing?

- Is the proper selection of gate types in order to trade-off power and timing requirements.

Power → Minimized
Frequency → Met

$$\begin{aligned} min \quad & Power(x) \\ s.t. \quad & Timing(x) \leq T \end{aligned}$$

Power → Minimized
Frequency → Met

$$\begin{aligned} min \quad & Timing(x) \\ s.t. \quad & Power(x) \leq P \end{aligned}$$

where *x* is the vector with the selected cell type for each cell

Gate Sizing

- Discrete
  - Also known as gate selection
  - NP-Complete
  - No exact solution can be found in reasonable time (general case)
- Continuous
  - Optimum solutions can be found in feasible time but relying on simplified delay models.

Problem Formulation – ISPD Contest 2012

$$\begin{aligned} min \quad & Leakage(x) \\ s.t. \quad & Timing(x) \leq T \end{aligned}$$

ISPD Contest 2013

International Symposium on Physical Design

- Static power (leakage) is the main concern.
- Static Timing Analysis used as measure.
- Simplified wire model: lumped capacitance.

Leakage Power

Power consumption when circuit is on but not “working”. Gets worse as dimensions shrinks

Lagrangian Relaxation

$$\begin{aligned} min \quad & Leakage(x) \\ s.t. \quad & Timing(x) \leq T \end{aligned}$$

Very hard to be solved!

Lagrangian Relaxation
Hard constraints become part of the objective function.

$$min \quad Leakage(x) + \lambda(Timing(x) - T)$$

A little bit easier...
heuristics still needed.

Lagrangian Multipliers

Problem Formulation – KKT Conditions

- Due to KKT conditions to optimality, we got...

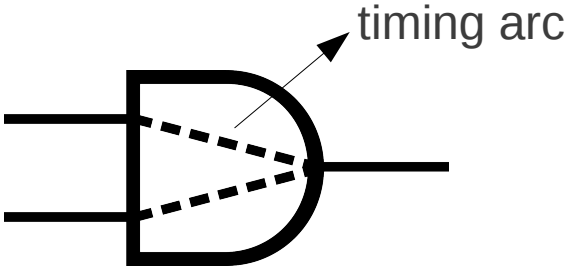
$$min \quad Leakage(x) + \lambda(Timing(x) - T)$$

equivalent

C. P. Chen, C. C.-N. Chu, and D. F. Wong
Fast and exact simultaneous gate and wire sizing by lagrangian relaxation
IEEE Trans. on Computer-Aided Design

$$min \quad Leakage(x) + \lambda delay_k$$

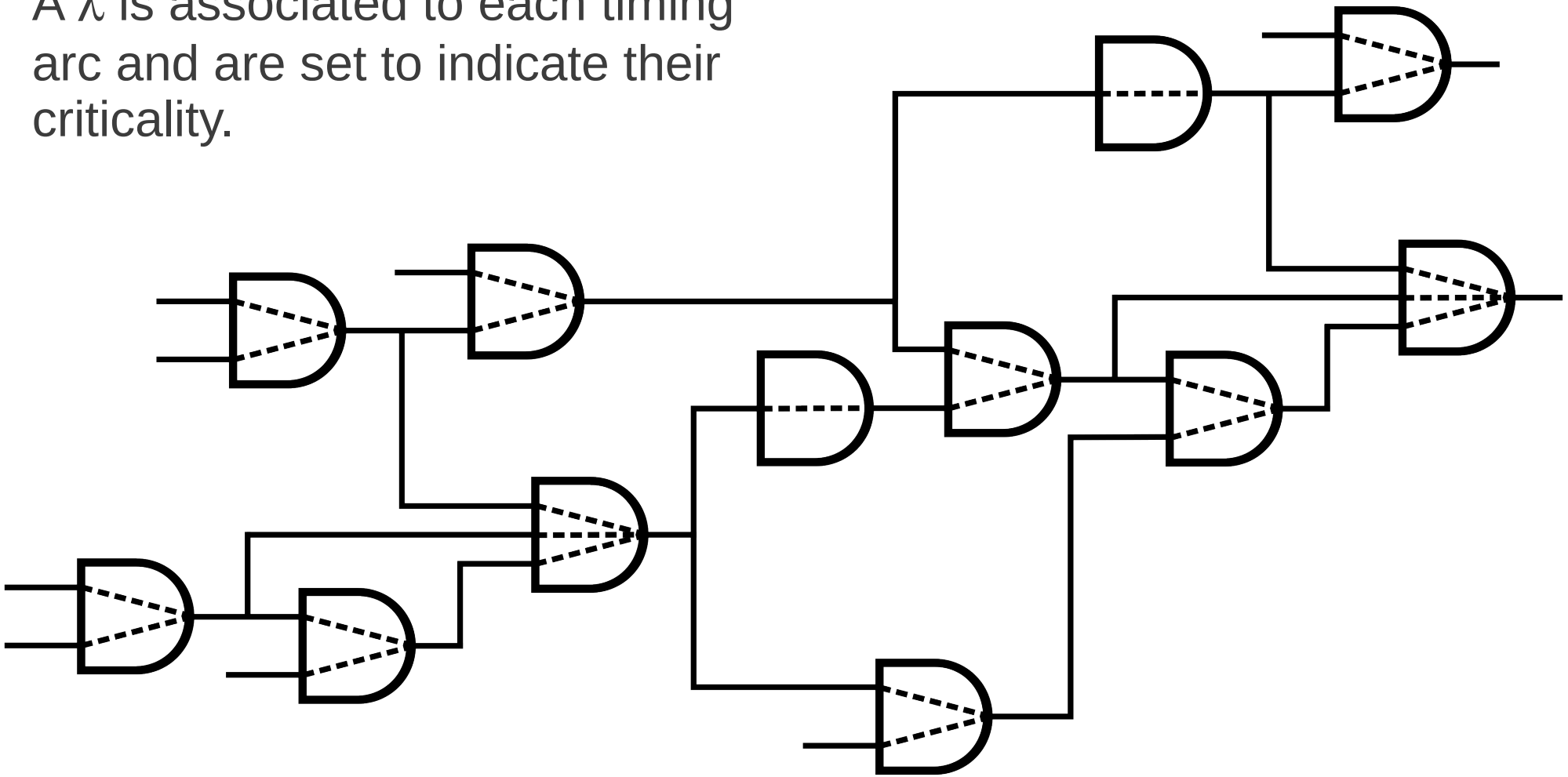
where $delay_k$ is the delay of timing ark k



Problem Formulation - After All

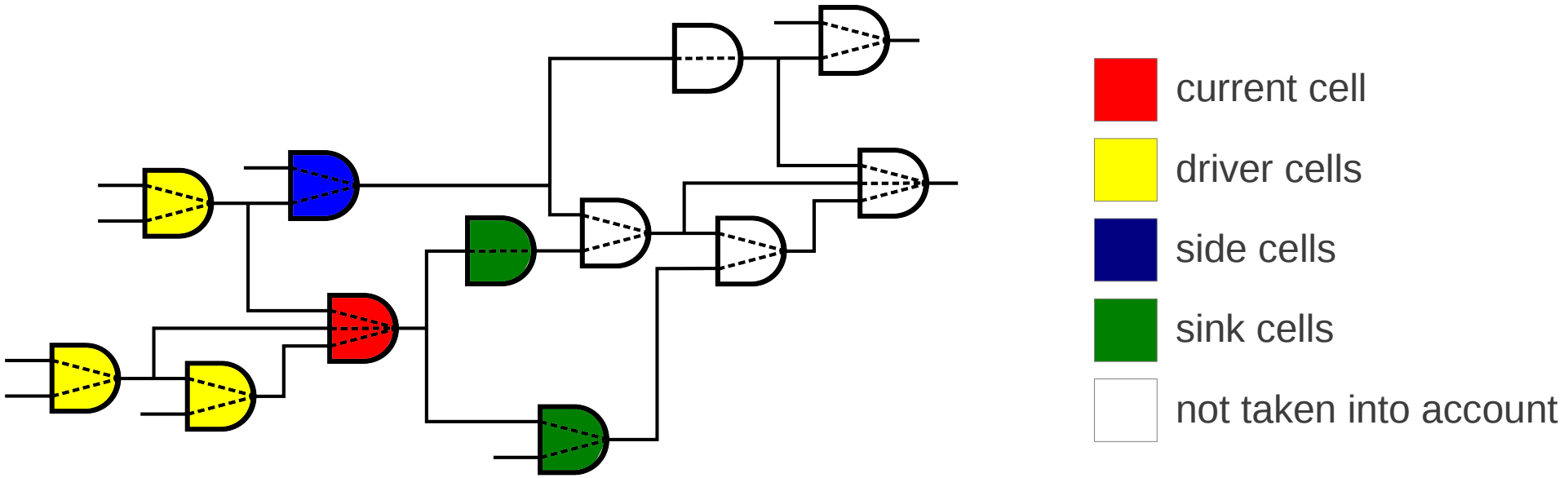
$$min \quad Leakage(x) + \lambda delay_k$$

A λ is associated to each timing arc and are set to indicate their criticality.



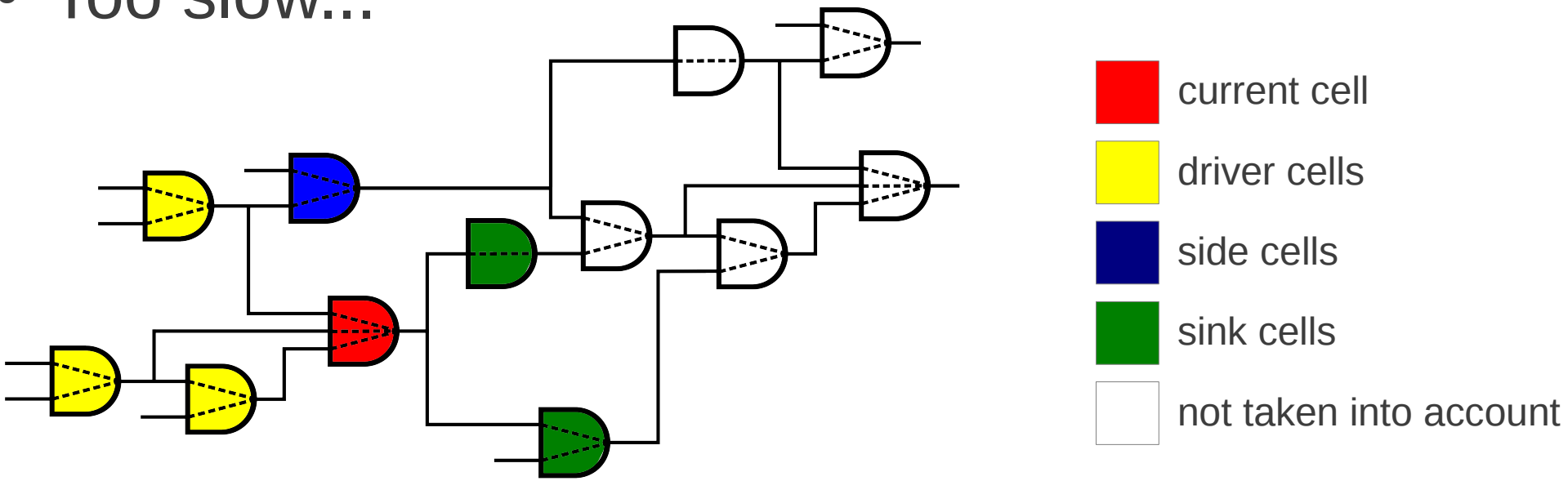
Approach

- $min \quad Leakage(x) + \lambda delay_k$
- for each cell in topological order
  - **try every cell type** and chose the one which **locally** best **trades-off leakage and delay** .



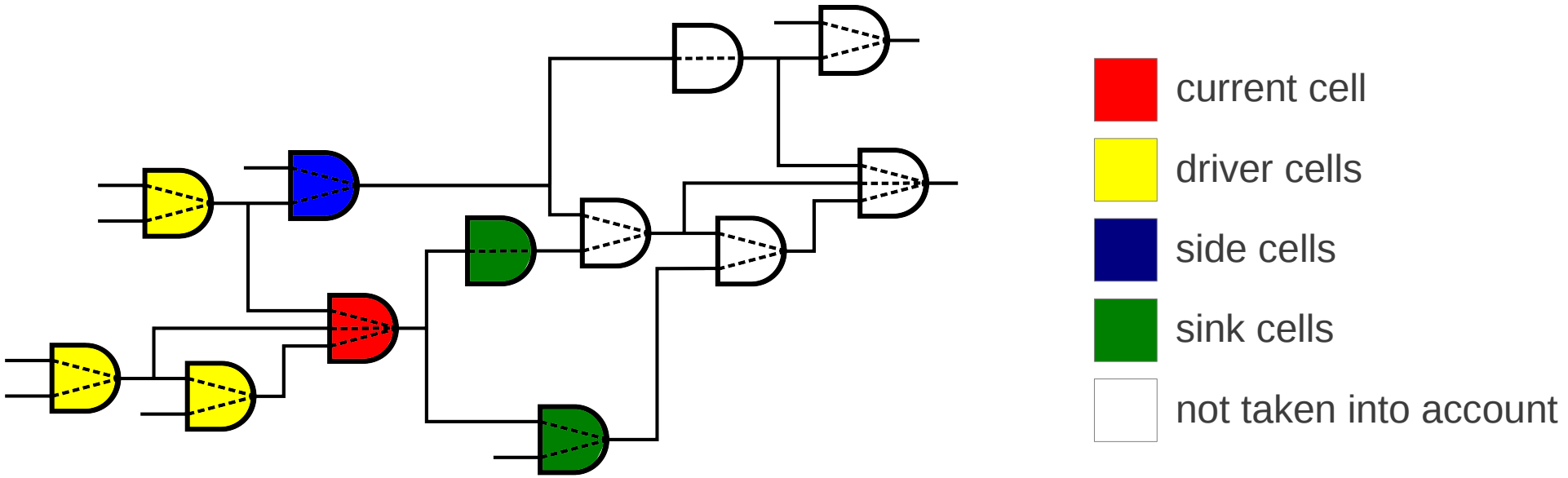
Why only local effects are taken into account?

- Well, a change in a cell type may affect several timing arc delays.
- But an incremental STA would be required for each cell change to keep timing updated.
- Too slow...

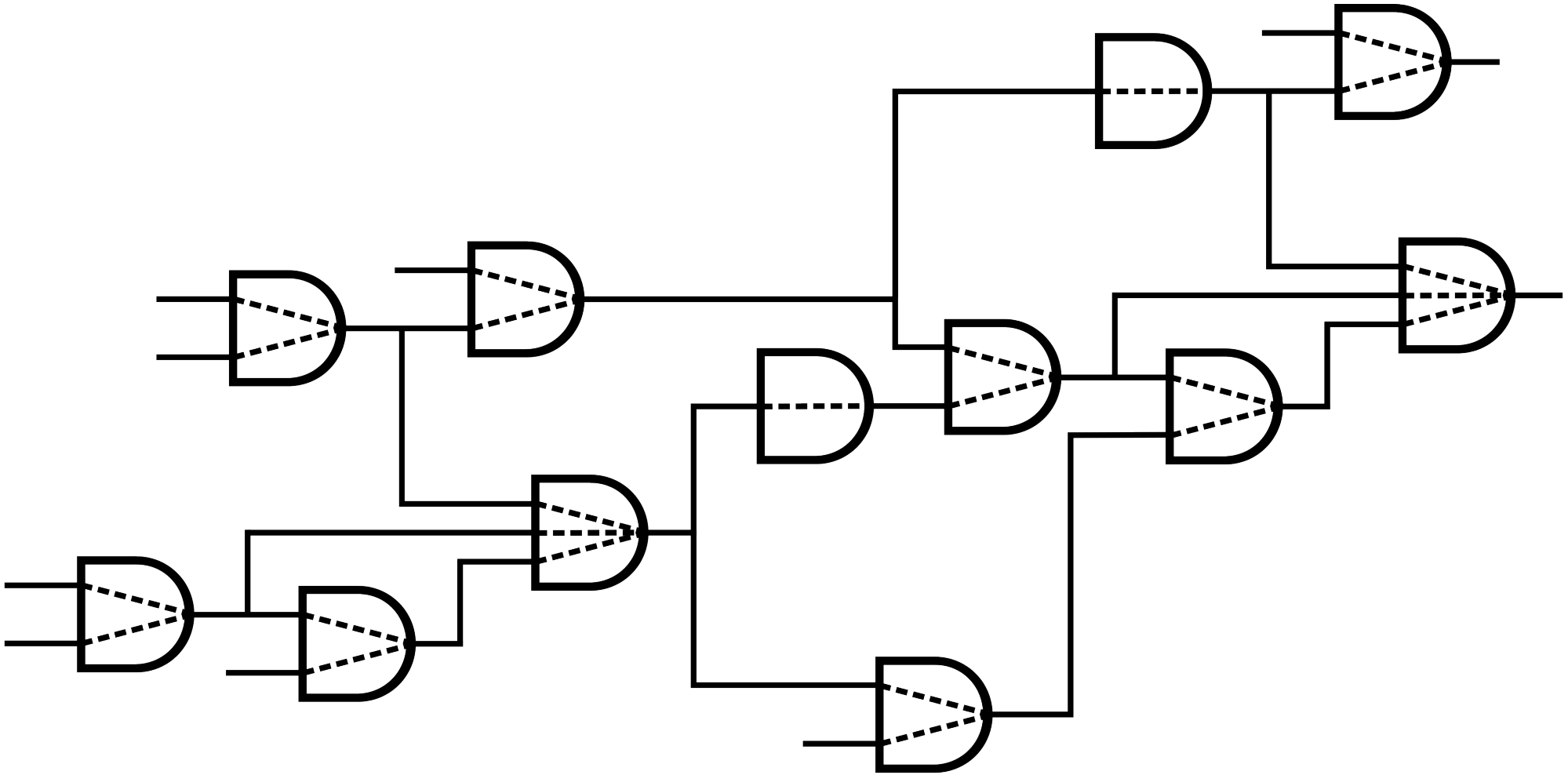


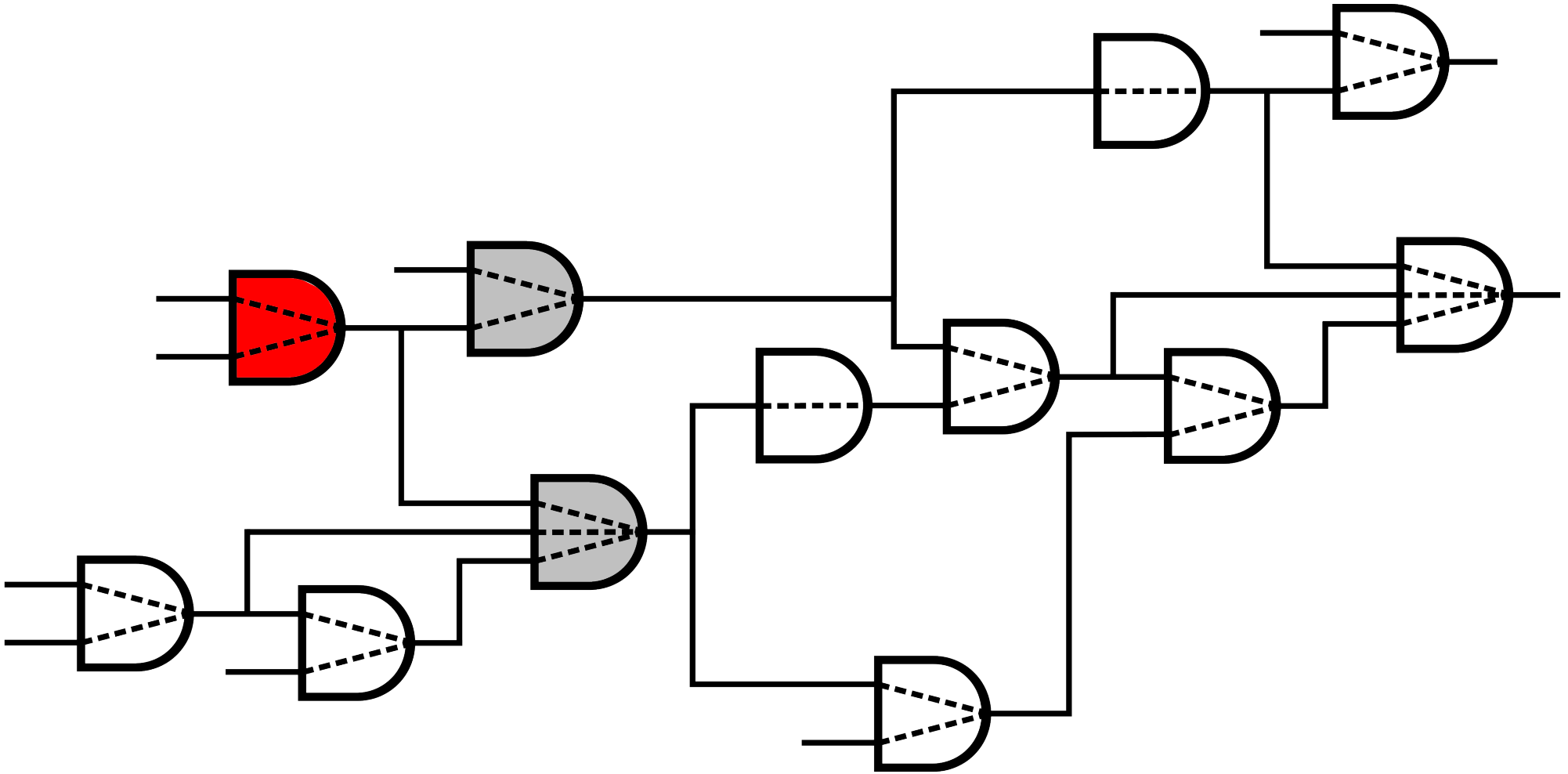
But isn't that too bad?

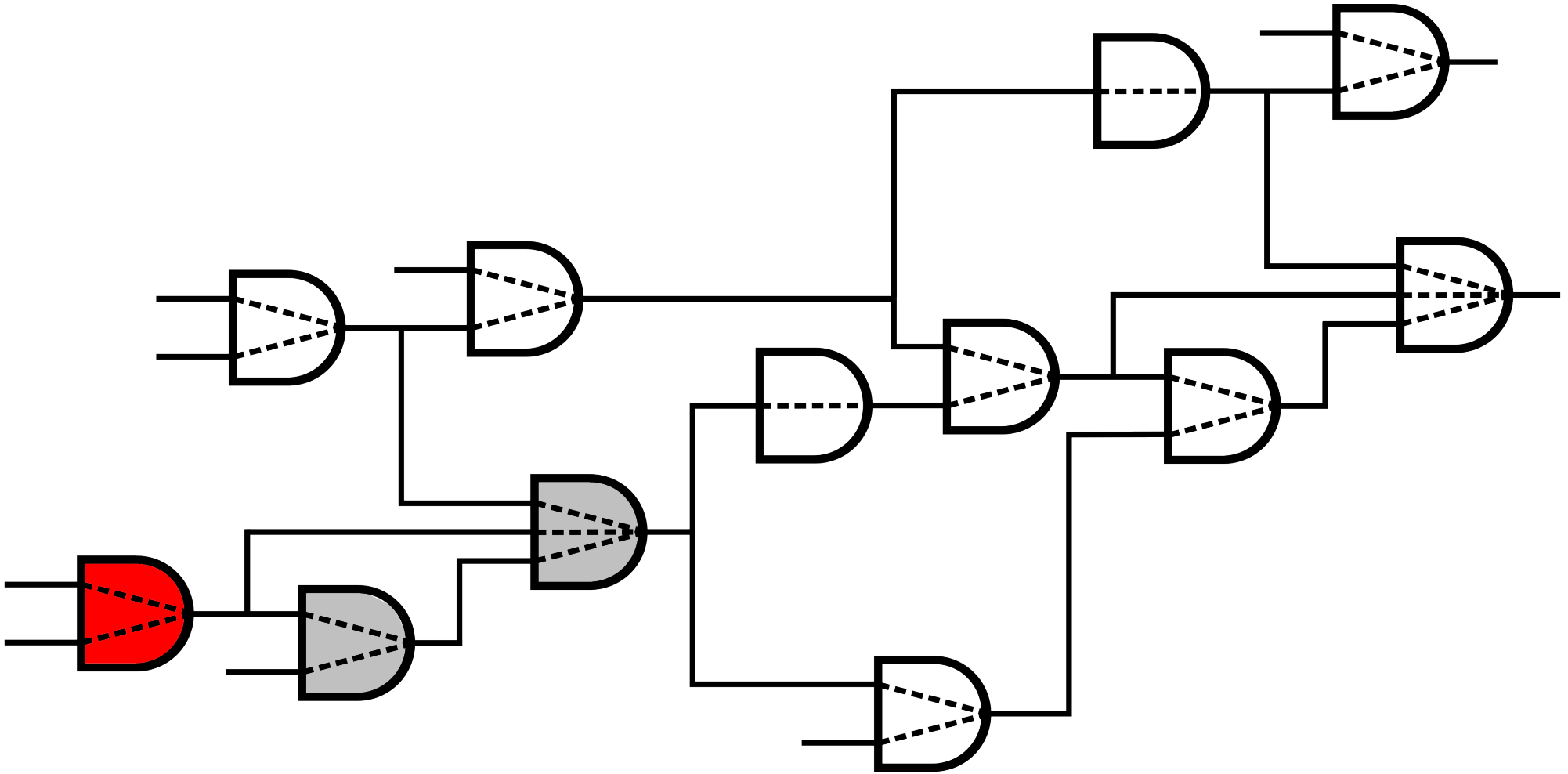
- Not totally, since the most effect occurs on the first levels.



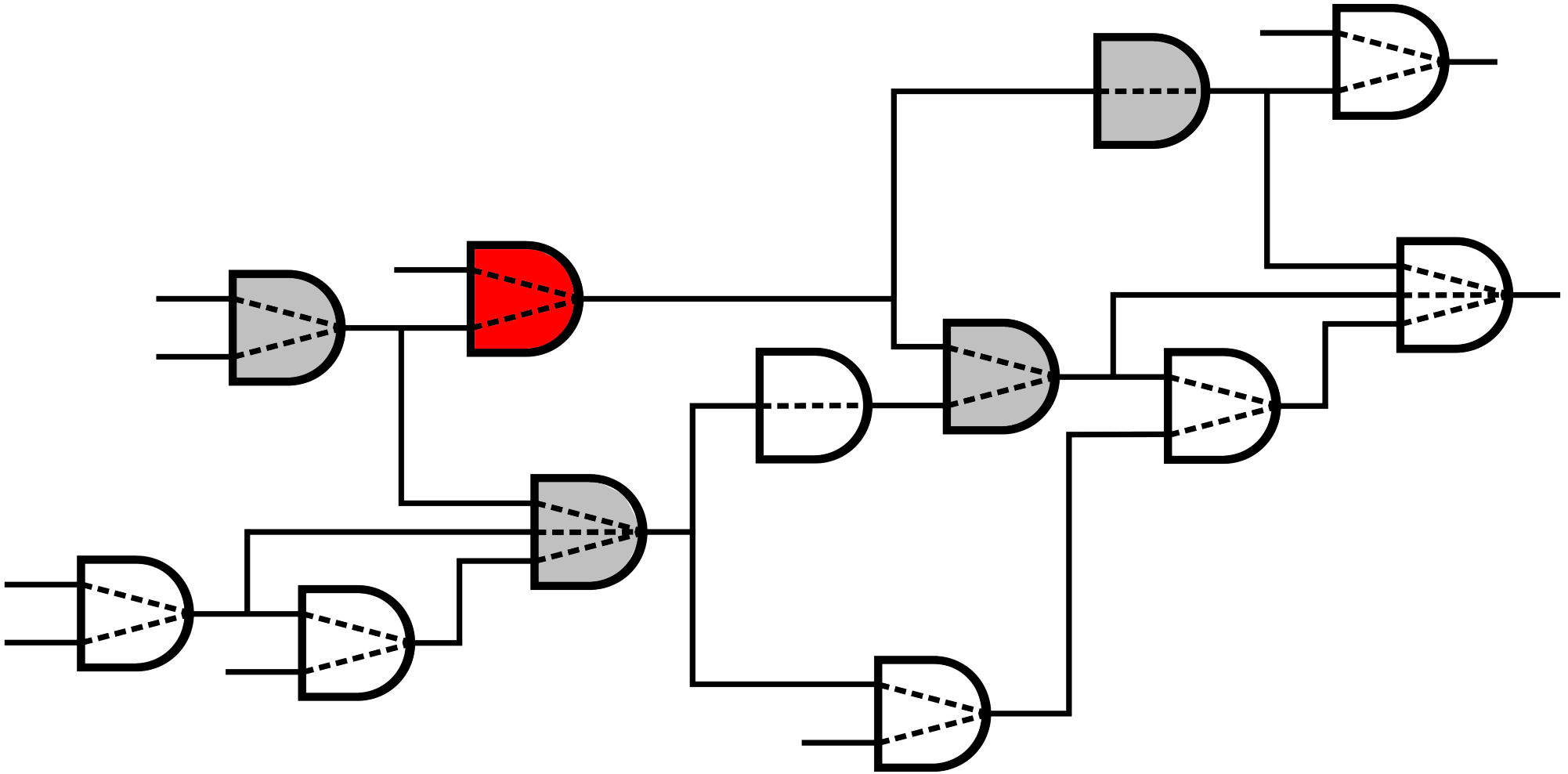
Greedy Algorithm



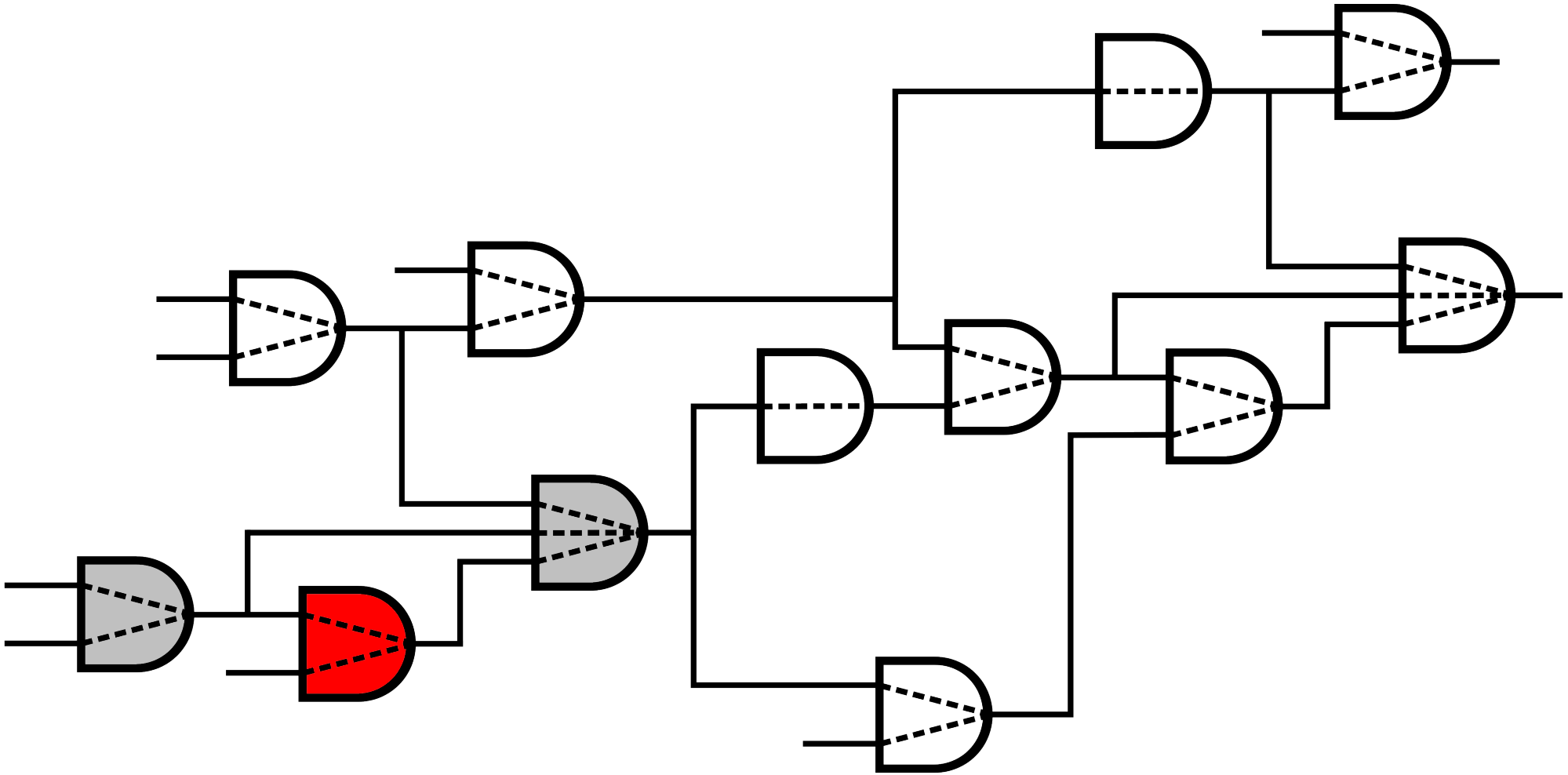




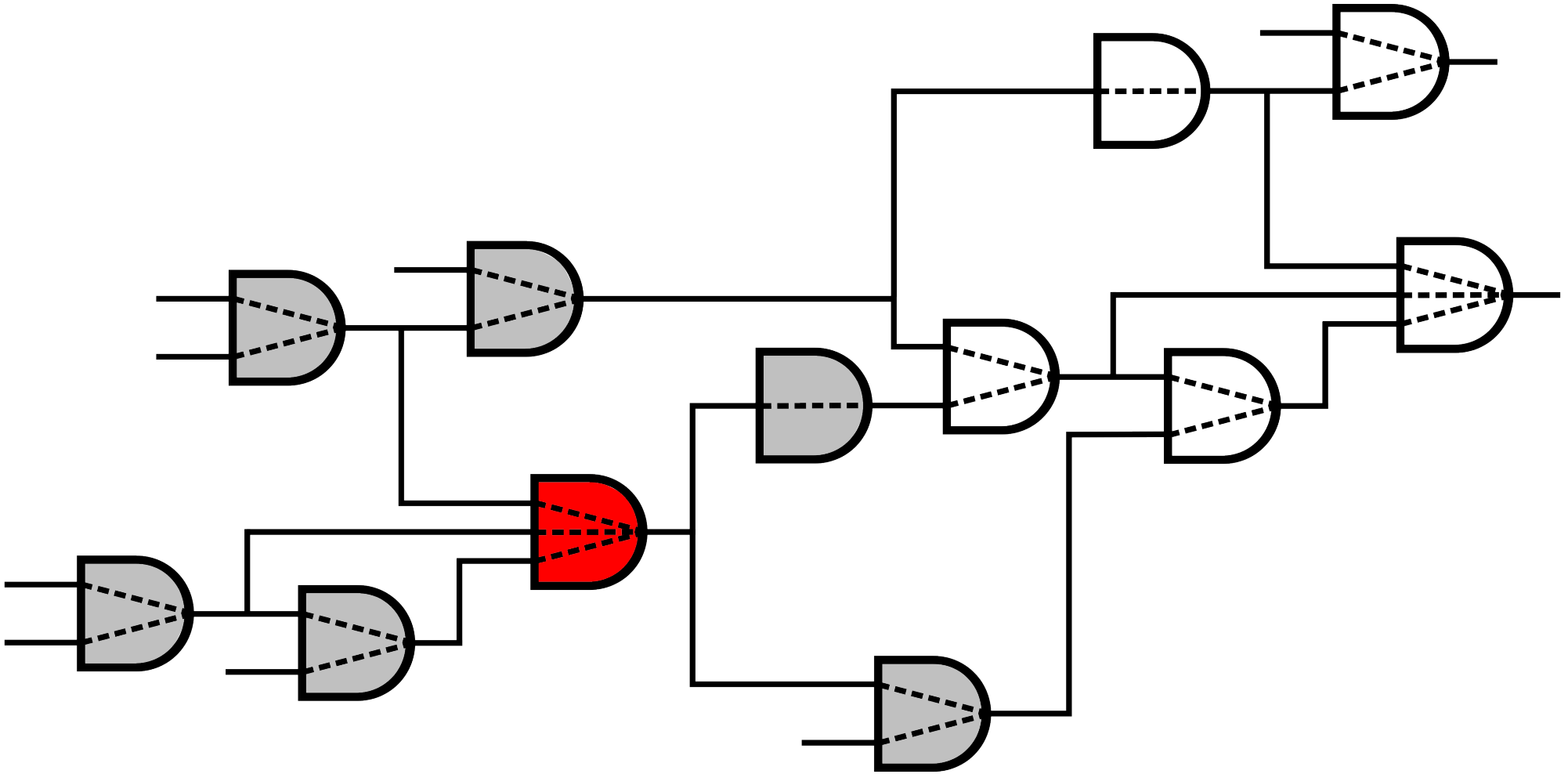
Greedy Algorithm



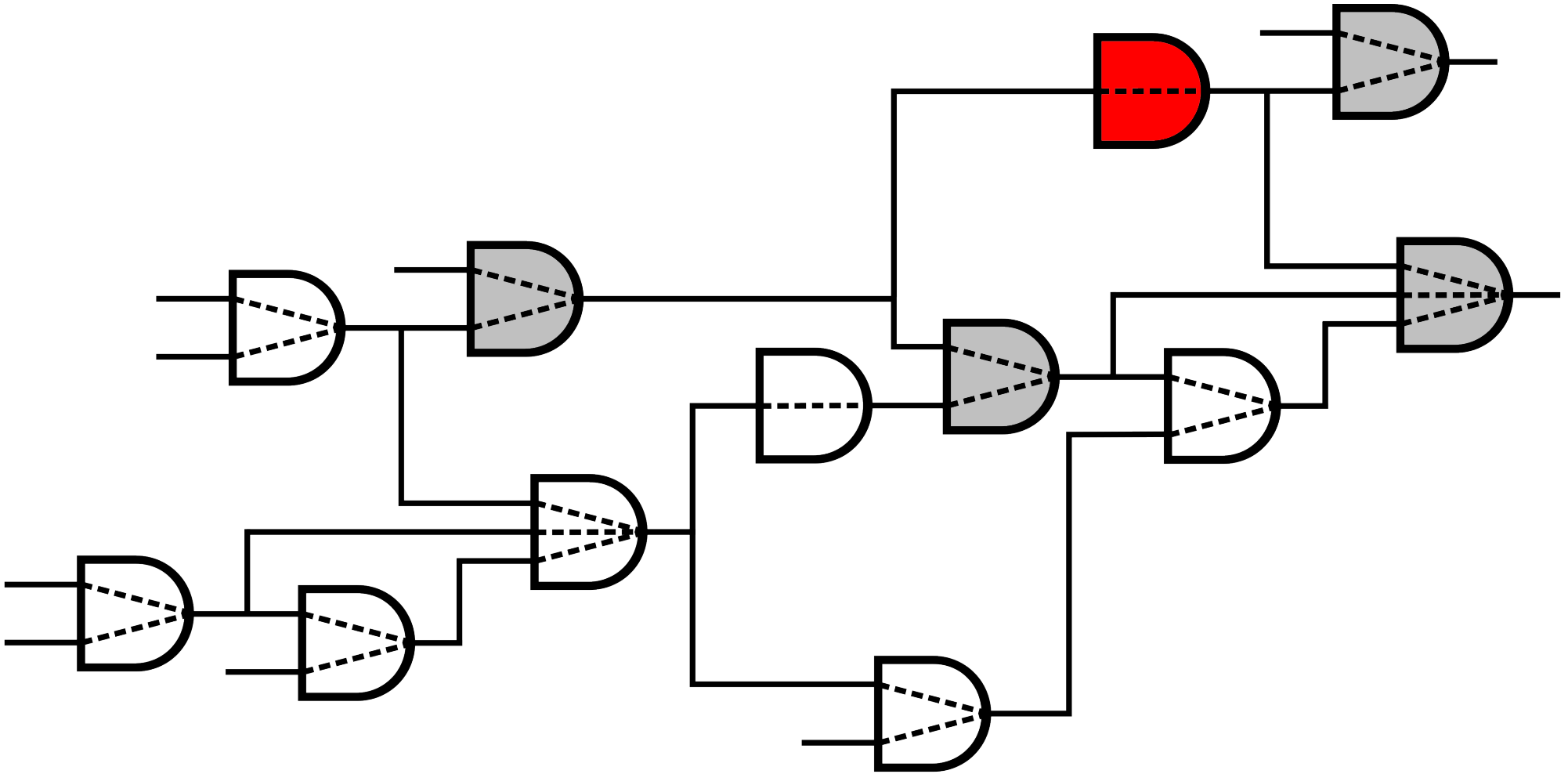
Greedy Algorithm



Greedy Algoritm

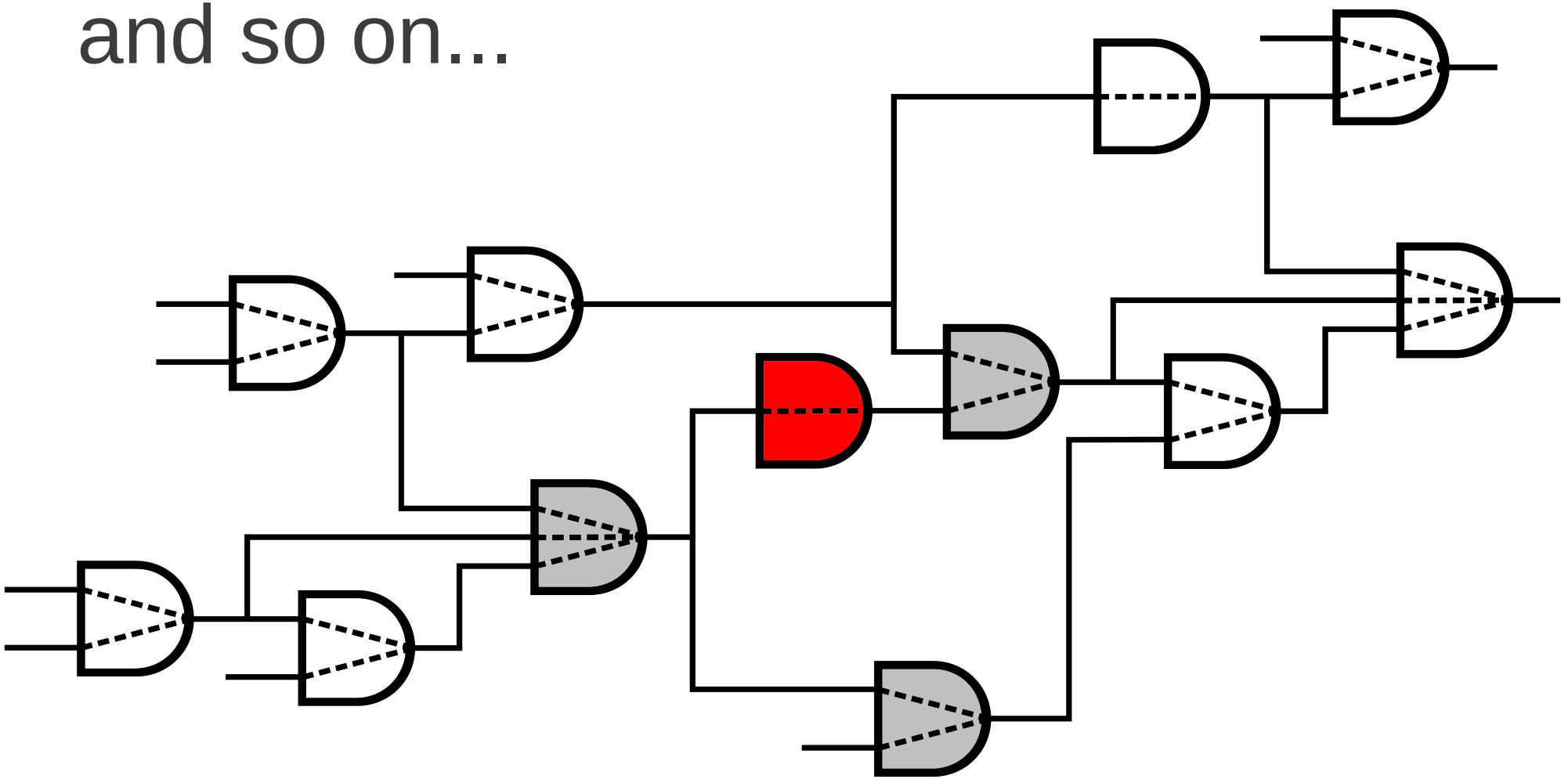


Greedy Algorithm



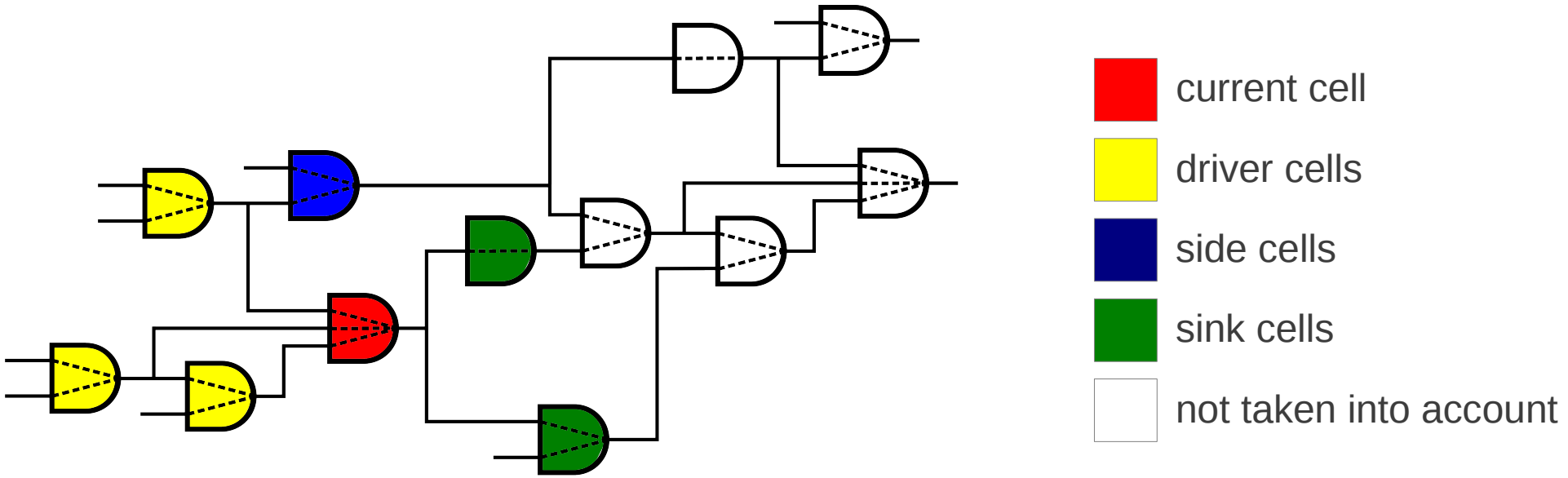
Greedy Algorithm

and so on...



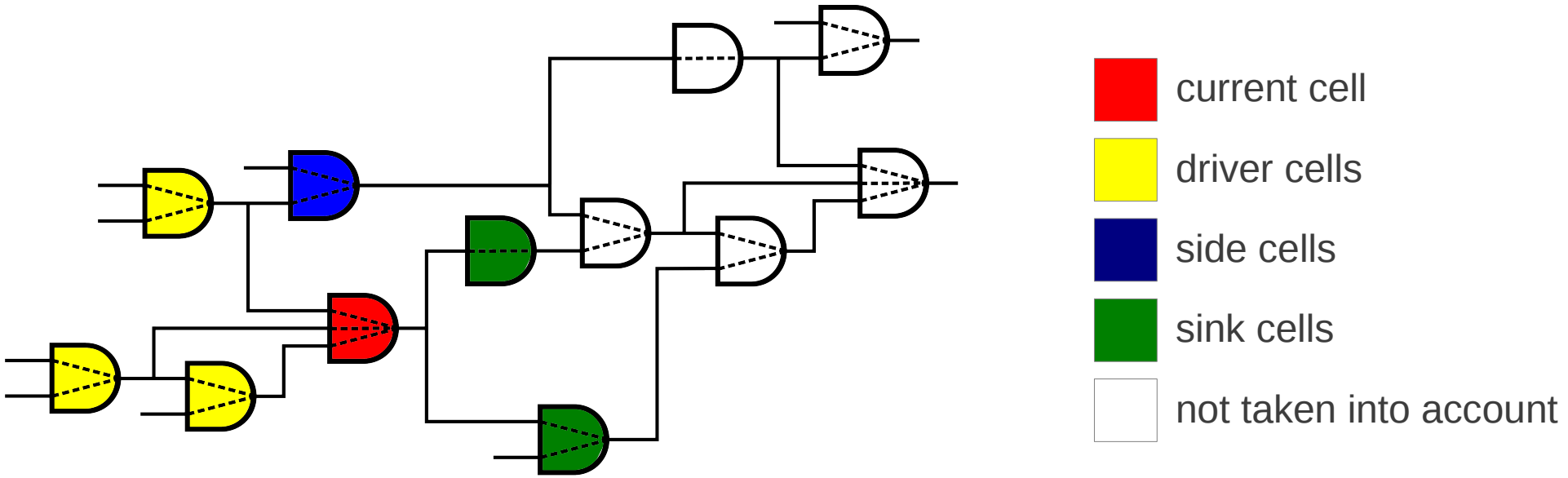
Can we do better?

- Dealing with the cells other than that in the vicinity.
- Propagate back the timing arc delays sensitivities.



Timing Arc Delay Sensitivities

- Idea: how a change in the middle of circuit affect the overall circuit time
  - without requiring to execute STA

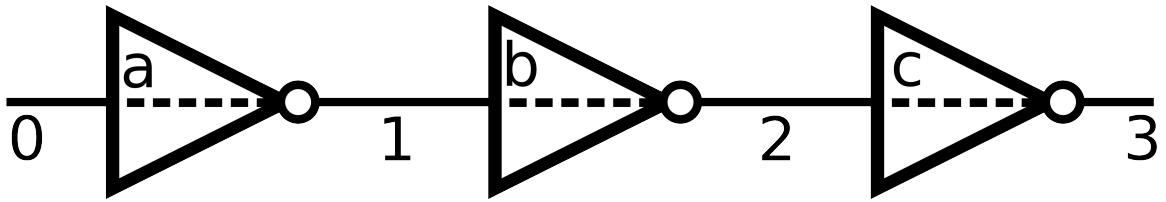


Timing Arc Delay Sensitivities

- for timing arc in reverse topological order
  - compute and propagate its sensitivity in delay due to a change in its input slew

$$\Delta delay_a = \Delta islew_0 \left[\frac{\delta delay_a}{\delta islew_a} + \frac{\delta oslew_a}{\delta islew_a} \left(\frac{\delta delay_d}{\delta islew_d} + \frac{\delta oslew_b}{\delta islew_b} \frac{\delta delay_c}{\delta islew_c} \right) \right]$$

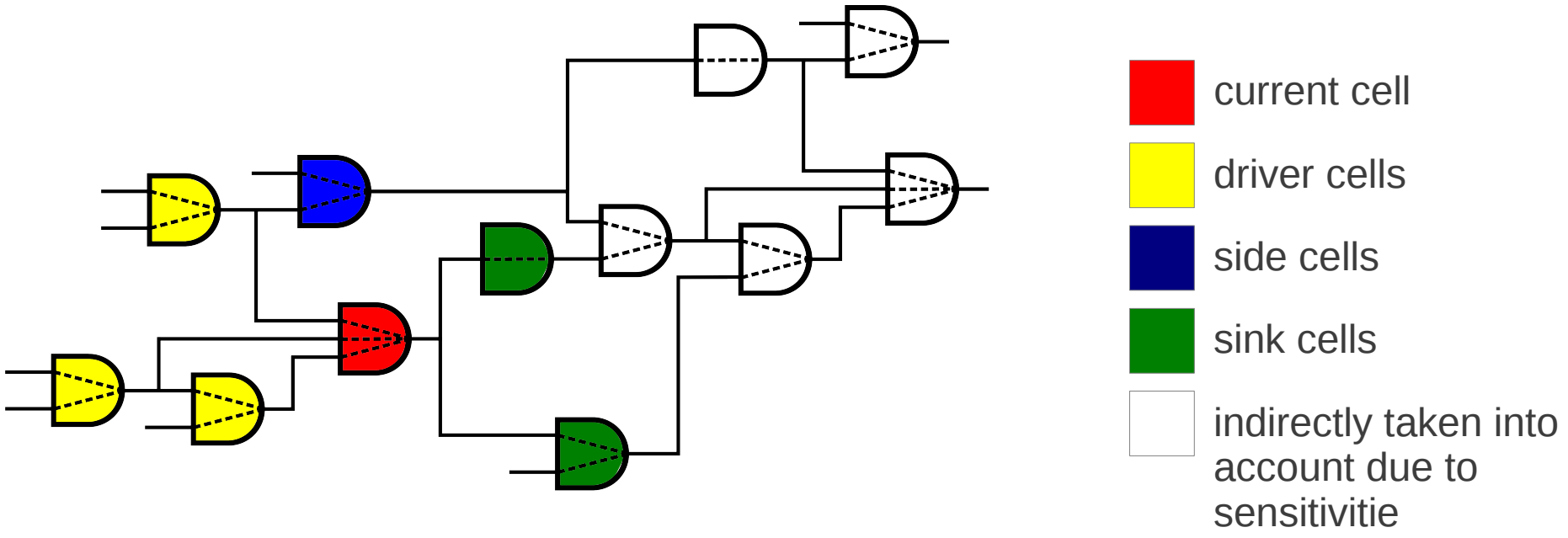
$$\Delta delay_c = \Delta islew_c \frac{\delta delay_c}{\delta islew_c}$$



$$\Delta delay_b = \Delta islew_1 \frac{\delta delay_b}{\delta islew_b} + \Delta delay_c$$

Timing Arc Delay Sensitivities

- Precise timing information is still only locally computed...
  - but sensitivities provides a way to see what may happen globally;
  - an no incremental STA required.



Power and Timing Recovery

- After selecting cells via Lagrangian:
  - sweep all cells trying to fix any timing violation left
    - usually our Lagrangian model provides a small timing violation
  - sweep all cells trying to decrease leakage power
- In these methods, an incremental STA is used for each cell change.

Results

- Smallest leakage power for every ISPD 2012 benchmark.
  - 23k cells circuit sized in less than 1min.
  - 861K cells circuit sized in less than 50min.

Benchmark	# of Comb. Cells	Leakage Power (W)			Power Saved (%)		Runtime (m)			OurLPRS Reduction (%)	
		Hu [1]	Li [2]	Our-LPRS	Compared to [1]	Compared to [2]	Hu [1]	Li [2]	Our-LPRS	Compared to [1]	Compared to [2]
DMA_slow	23K	0.145	0.153	0.132	8.73	13.50	9.90	0.60	0.79	92.02	-31.67
DMA_fast		0.299	0.281	0.238	20.29	15.19	13.90	0.60	0.92	93.38	-53.33
pci_bridge32_slow	30K	0.111	0.111	0.096	13.31	13.31	10.20	1.20	0.87	91.47	27.50
pci_bridge32_fast		0.183	0.167	0.136	25.51	18.37	13.00	1.20	0.92	92.92	23.33
des_perf_slow	102K	0.614	0.671	0.570	7.14	15.03	70.10	6.00	25.31	63.89	-321.83
des_perf_fast		1.842	1.93	1.395	24.27	27.73	82.70	6.60	16.37	80.21	-148.03
vga_lcd_slow	148K	0.351	0.375	0.328	6.61	12.59	87.50	7.80	5.67	93.52	27.31
vga_lcd_fast		0.471	0.46	0.413	12.22	10.12	45.60	10.20	8.37	81.64	17.94
b19_slow	213K	0.583	0.604	0.564	3.28	6.64	213.90	10.20	9.15	95.72	10.29
b19_fast		0.771	0.784	0.717	7.06	8.61	206.50	12.00	11.75	94.31	2.08
leon3mp_slow	540K	1.341	1.4	1.334	0.53	4.72	1,274.00	43.80	38.98	96.94	11.00
leon3mp_fast		1.487	1.64	1.443	2.99	12.04	1,323.20	54.60	46.62	96.48	14.62
netcard_slow	861K	1.77	1.78	1.763	0.41	0.97	299.90	48.00	34.39	88.53	28.35
netcard_fast		1.861	2.18	1.841	1.07	15.55	1,096.90	88.80	47.41	95.68	46.61
Avg.	-	**0.845**	**0.895**	**0.784**	**9.53**	**12.45**	-	-	-	-	-
Sum (h)	-	-	-	-	-	-	**79.12**	**4.86**	**4.13**	**94.79**	**15.12**

Conclusions

- Lagrangian Relaxation fits well to model the discrete gate sizing problem.
  - Mainly when simplified by KKT conditions
- A change in a cell type may affect almost the whole circuit.
  - However the most effects are nearby the changed cell.
  - The remaining effect can be estimated propagating back the delay sensitivities of timing arcs.

EMICRO/SIM 2013
XV Escola de Microeletrônica Sul / 28º Simpósio Sul de Microeletrônica
Porto Alegre, 29 de abril a 3 de maio de 2013

Simultaneous Gate Sizing and Vth Assignment using Lagrangian Relaxation and Delay Sensitivities

Guilherme Flach, Tiago Reimann, Gracieli Posser
Marcelo Johann, Ricardo Reis

gaflach@inf.ufrgs.br

UFRGS
PGMICRO
PROGRAMA DE
PÓS-GRADUAÇÃO EM MICROELETRÔNICA