

# SUPERCONDUCTIVE SINGLE FLUX QUANTUM LOGIC DEVICES AND CIRCUITS: STATUS, CHALLENGES, AND OPPORTUNITIES

---

**M. Pedram**

USC ECE

pedram@usc.edu

<http://coldflux.usc.edu/>

<https://discoverexpedition.usc.edu/>

Talk given at ISLPED

Aug. 2, 2022

# Going Faster with Less Energy

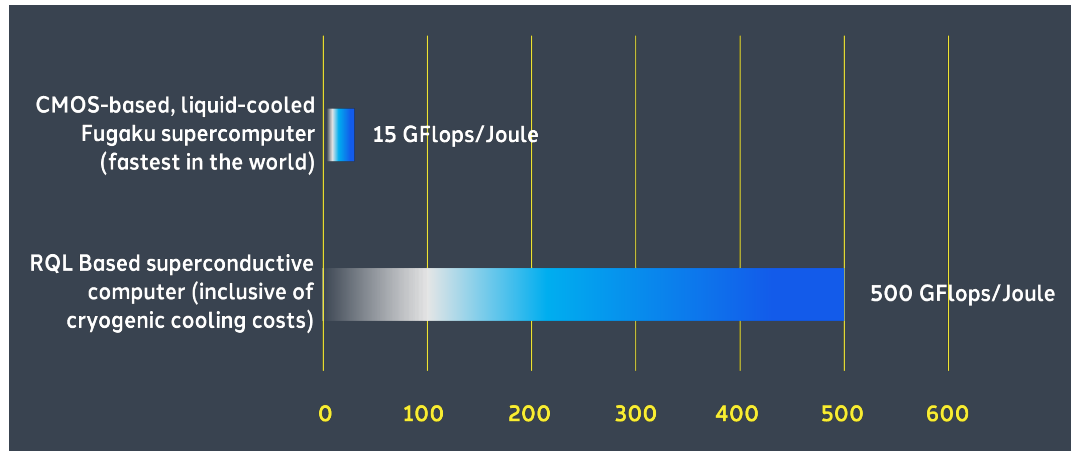
- Very high clock frequencies
  - High single-threaded performance through short switching times and fast signal propagation
- Ultra low switching energy
  - Energy related to the generation of the bit information should be made as small as possible

**Superconductor circuits meet both of the aforesaid requirements and have become a strong candidate for a beyond-CMOS computing fabric. This technology is known for the high clock frequencies, 2–50 GHz, and the low energy dissipation per logical operation, down to several zJ [1].**

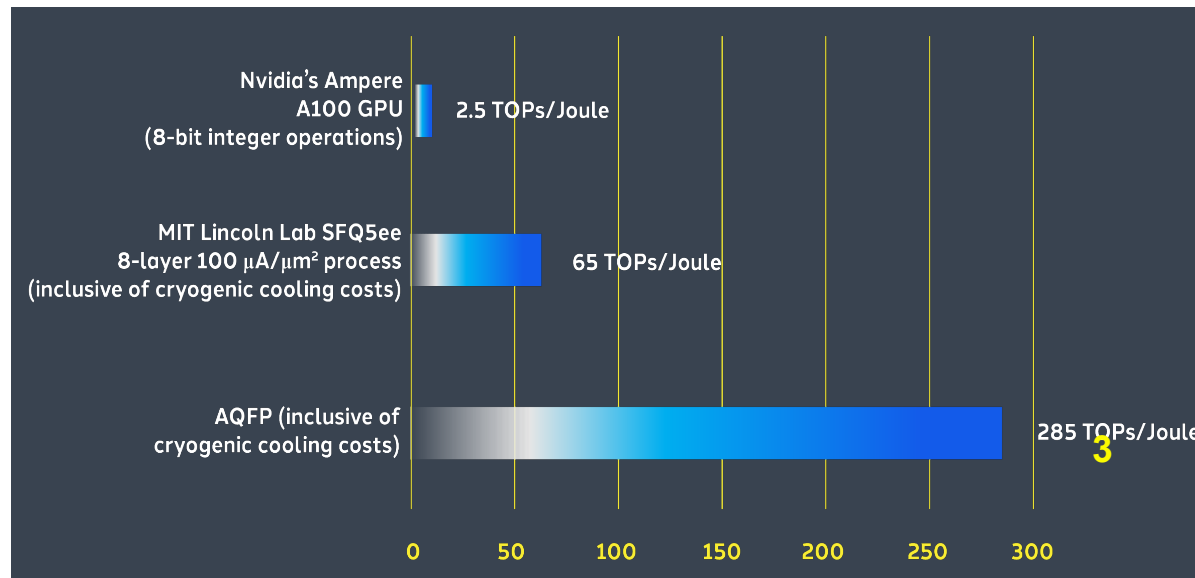
[1] K. K. Likharev and V. K. Semenov, RSFQ logic/memory family: A new Josephson-junction technology for sub-terahertz-clock-frequency digital systems, *IEEE Trans. Appl. Supercond.* 1, 3 (1991).

[2] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, Measurement of 10zJ energy dissipation of adiabatic quantum flux-parametron logic using a superconducting resonator, *Appl. Phys. Lett.* 102, 052602 (2013).

# The Case for Superconductor Circuits



System-level energy consumption (GFlops/Joule)



Neural network inference energy consumption (TOPs/Joule)

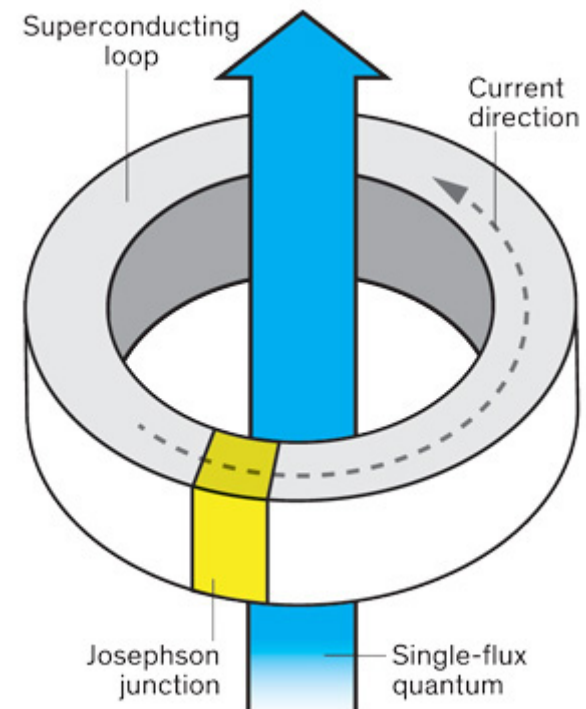
# Single Flux Quantum Logic

- Ginsburg-Landau theory: Superconductivity is a second-order phase transition, where **an ordered state emerges from a disordered state at low temperatures**
- Bardeen-Cooper-Schrieffer theory: Superconductivity is the **condensation of bound pairs of electrons**, where these Cooper pairs are connected through electron-phonon interactions
- Logic circuits **based on magnetic flux quantization** rather than conductance modulation
- Allows discrete representation of information in the form of Single Flux Quantum (SFQ)

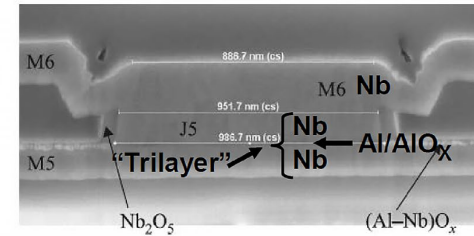
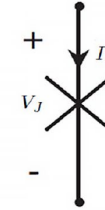
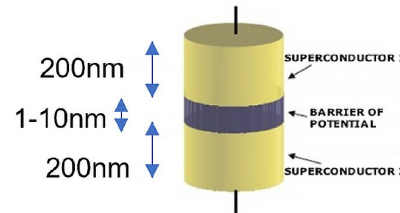
$\Phi_0 = \hbar/2e \approx 2.07 \times 10^{-15}$  volt-second (ampere-henry=weber)

$\hbar$  is the Planck constant,  $e$  is the electron charge

- Only an integer number of flux quanta can exist in a superconductive loop



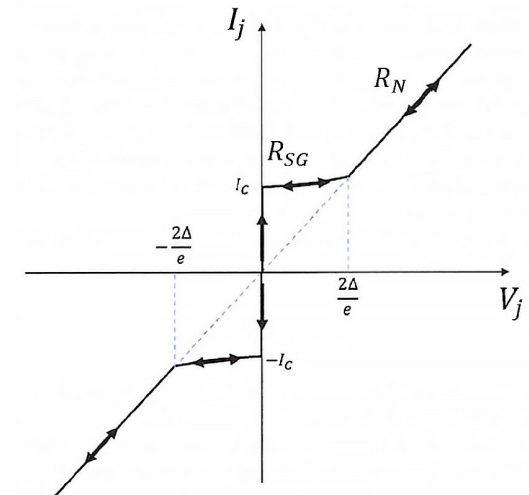
- **Josephson junction (JJ)** is the nonlinear switching element in superconductive circuits
- Consists of **two superconductors that are weakly coupled through a tunnel barrier (insulator)**



Most common junction for superconductor electronics is SIS planar junction: Nb/Al-AlO<sub>x</sub>/Nb @4.2K

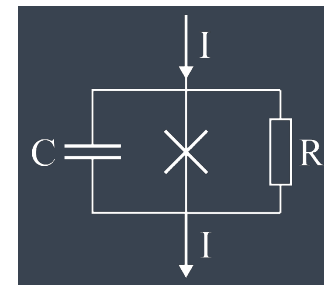
- Superconducting state phase difference  $\phi$  exists over JJ, producing supercurrent:

$I_j(t) = I_C \sin\phi(t)$  with  $d\phi/dt = 2\pi V_j(t)/\Phi_0$   
 where  $I_C$  denotes JJ critical current and  $\Phi_0 = \hbar/(2e)$   
 (Josephson equations define **current-phase relationship** and its evolution)



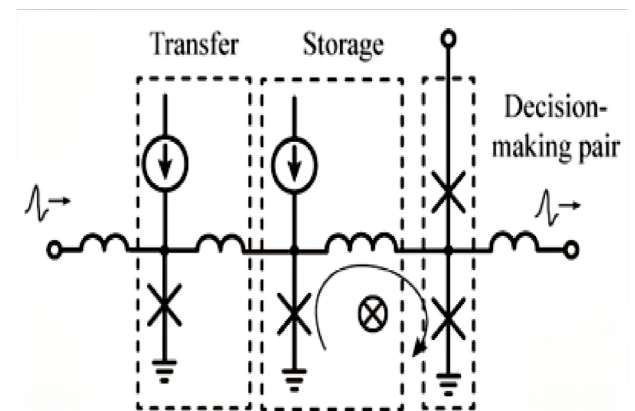
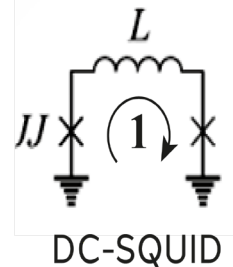
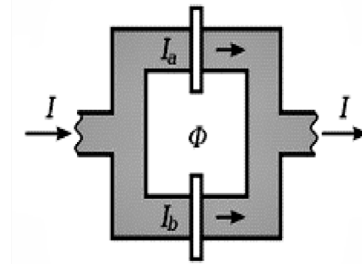
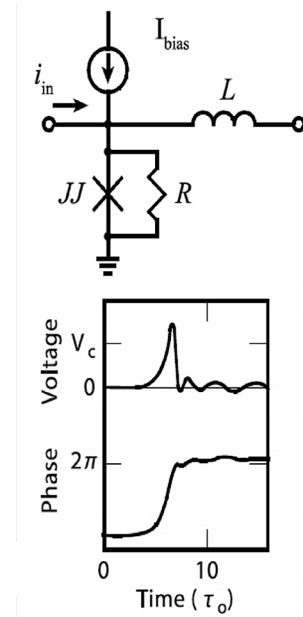
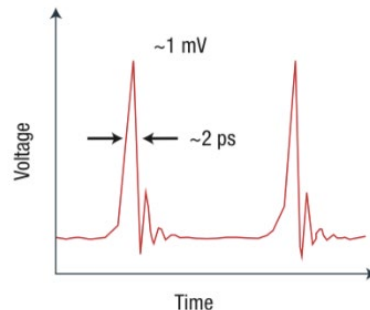
Overdamped JJ (small R and C), with Stewart-McCumber parameter  $\beta_C \approx 1$  is critically damped; this is the primary type of JJ

- **DC Josephson effect:** Constant DC current bias  $I_j < I_C$  results in constant phase difference  $\phi$  and zero voltage across JJ (tunneling of Cooper pairs through barrier produces supercurrent w/o applied voltage)
- **AC Josephson effect:** Constant voltage  $V_j$  applied across JJ results in linearly increasing phase difference and current oscillating at  $f = 2\pi V_j(t)/\Phi_0$



Resistively capacitively shunted junction (RCSJ) model

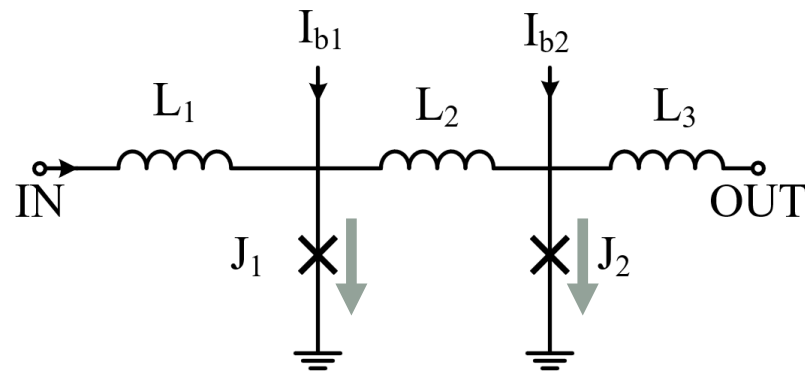
- Rapid SFQ (RSFQ), invented c. 1985 by O. Mukhanov, V. Semenov, and K. Likharev
- **Pulse-based logic family**, inherently synchronous, limited fanout drive
- Basic element of RSFQ logic is superconducting loop interrupted by one or more JJs
  - If the loop inductance  $L$  is high enough such that  $I_C L \geq \Phi_0$ , then SFQ can be held in the loop as a **persistent current representing logical one**, whereas an **absence of SFQ means logical zero**
  - Magnetic flux ejected from a superconducting loop through a JJ takes the form of **tiny voltage pulses (mV, ps)**
- Building blocks of SFQ circuits
  - **Transfer and storage sections**
  - **Decision-making pairs**



# Josephson Transmission Line

Josephson Transmission Line (JTL) shows how the SFQ devices transfer the information

The SQUID of JTL has one stable state. In this state, the current flow through both JJs is close to their critical current



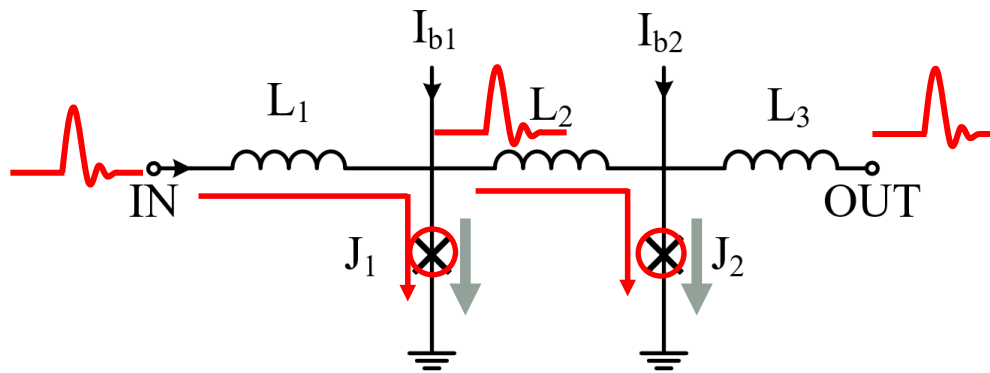
# Josephson Transmission Line



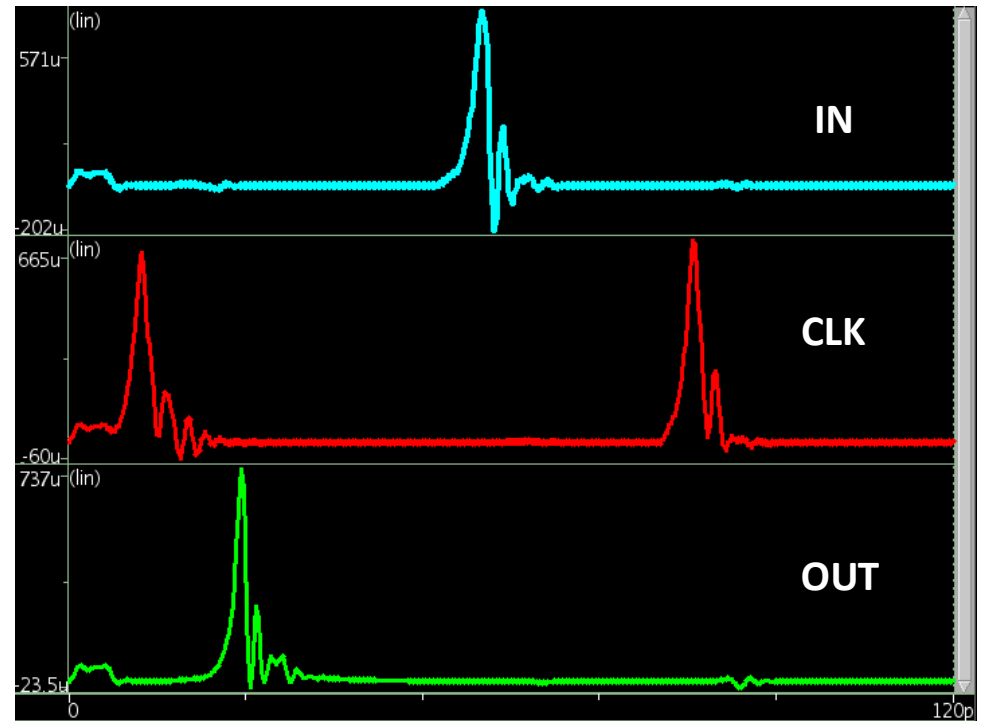
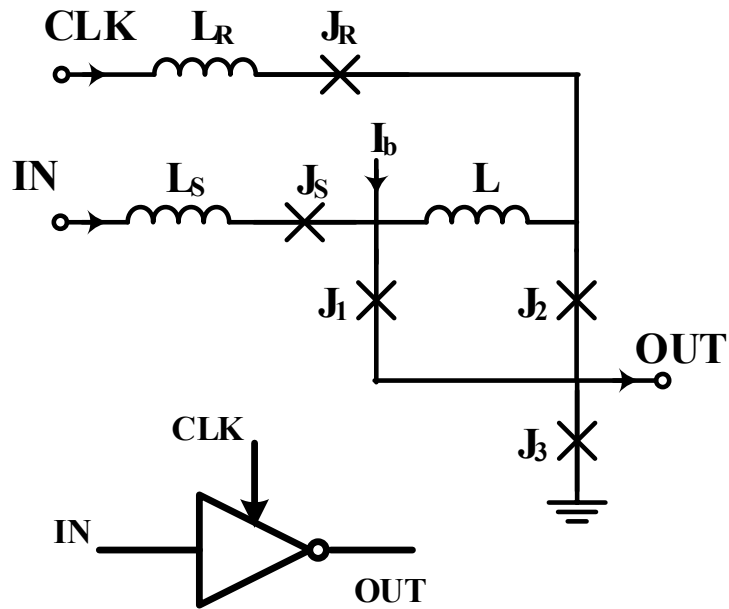
When we send an input voltage pulse to the JTL, the current flow through the  $J_1$  is increasing and it exceeds the critical current



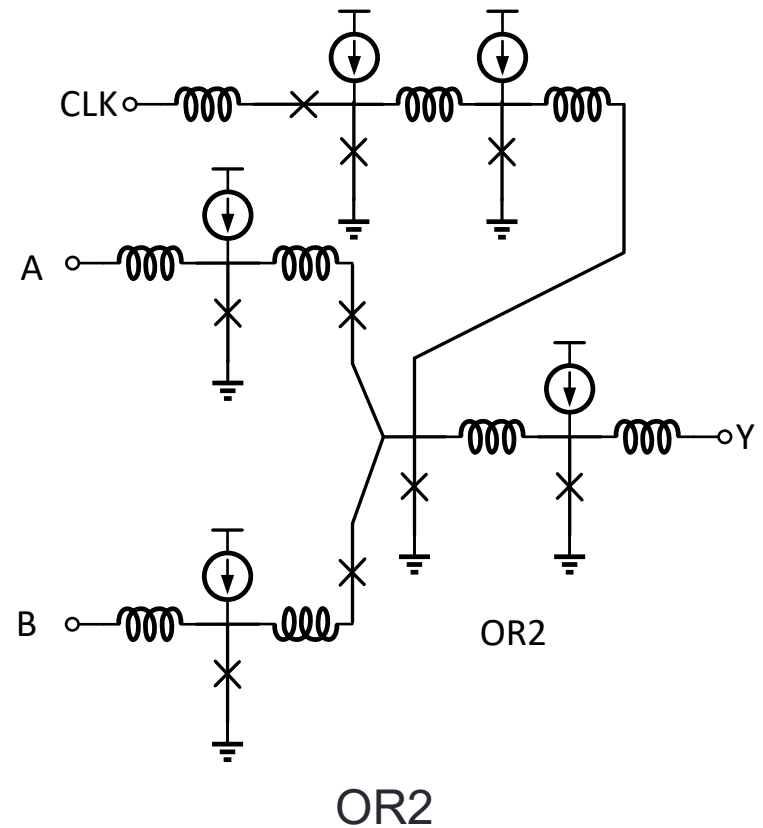
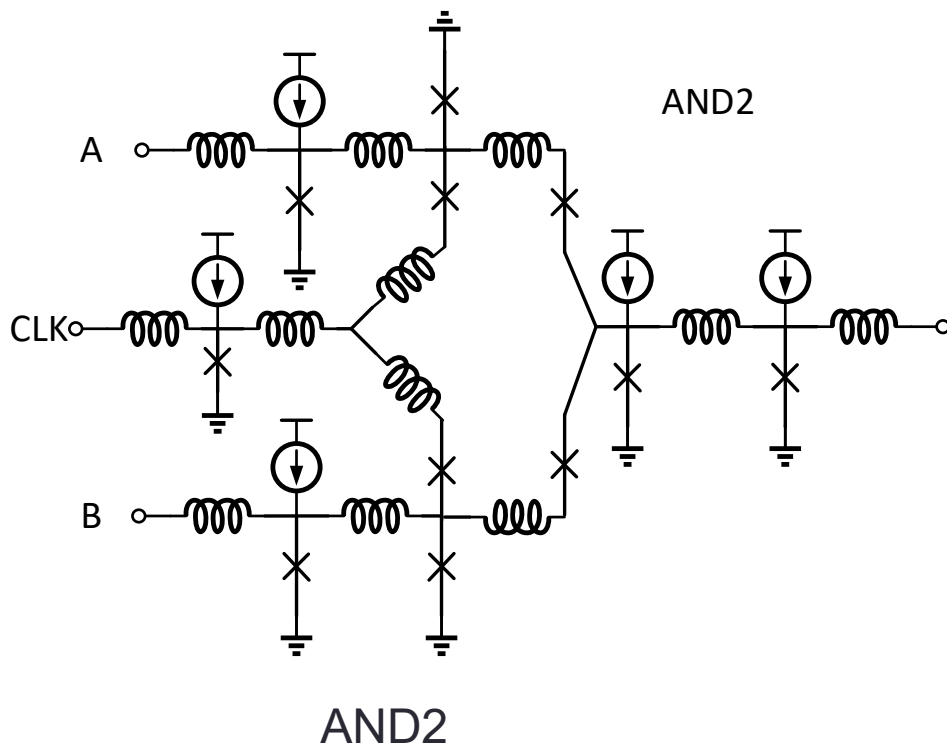
$J_1$  switches and creates a voltage pulse, the new pulse will increase the current flow through  $J_2$ , then  $J_2$  will switch and create an output voltage



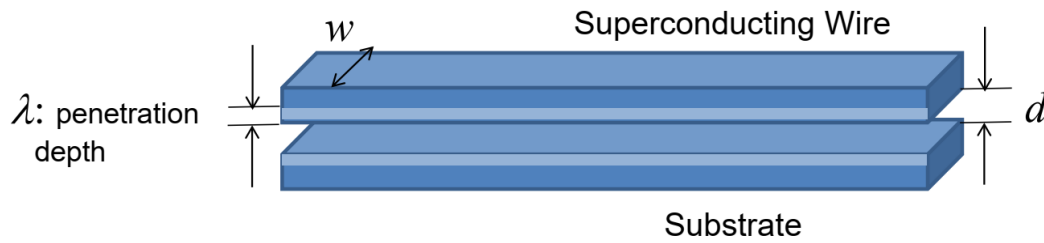
# RSFQ Inverter



# RSFQ Logic Gates

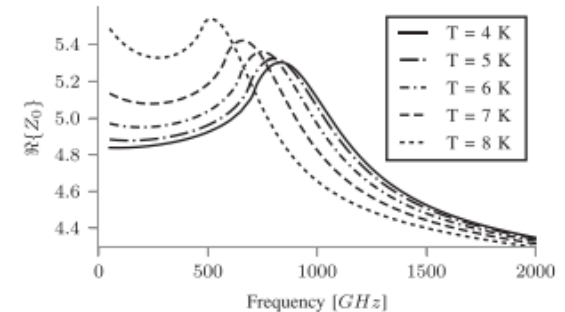
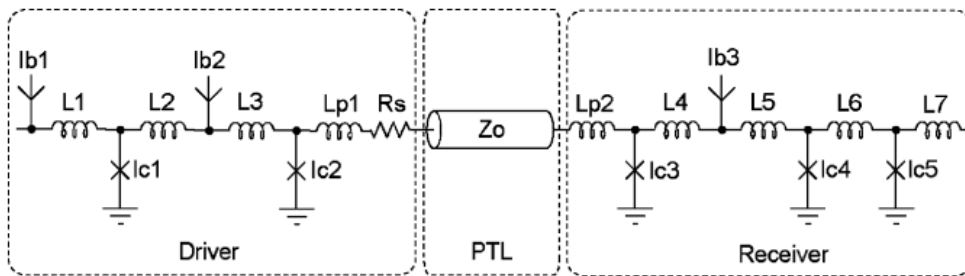


# Key Advantage: Passive Transmission Lines



$$v \cong \frac{1}{\sqrt{\epsilon\mu(1 + 2\lambda/d)}}$$

@ 100GHz, T=5K:  
 50 nH/m  
 100  $\Omega$ /m  
 0.34 speed of light



- Superconducting transmission is dispersion free, i.e., we have **ballistic propagation of an SFQ pulse** [5]
- Speeds up to 1/3 of the speed of light are achieved (**Typical delay: 10 ps/mm**)
- Needs PTL driver and receiver circuits, **Delay scales linearly with PTL length** [6].
- Must be impedance matched to the JJs of driver and receiver circuits. Junctions in the latest fabrication processes allow about 5 to 10  $\Omega$  lines, which reduces line width for tighter integration.

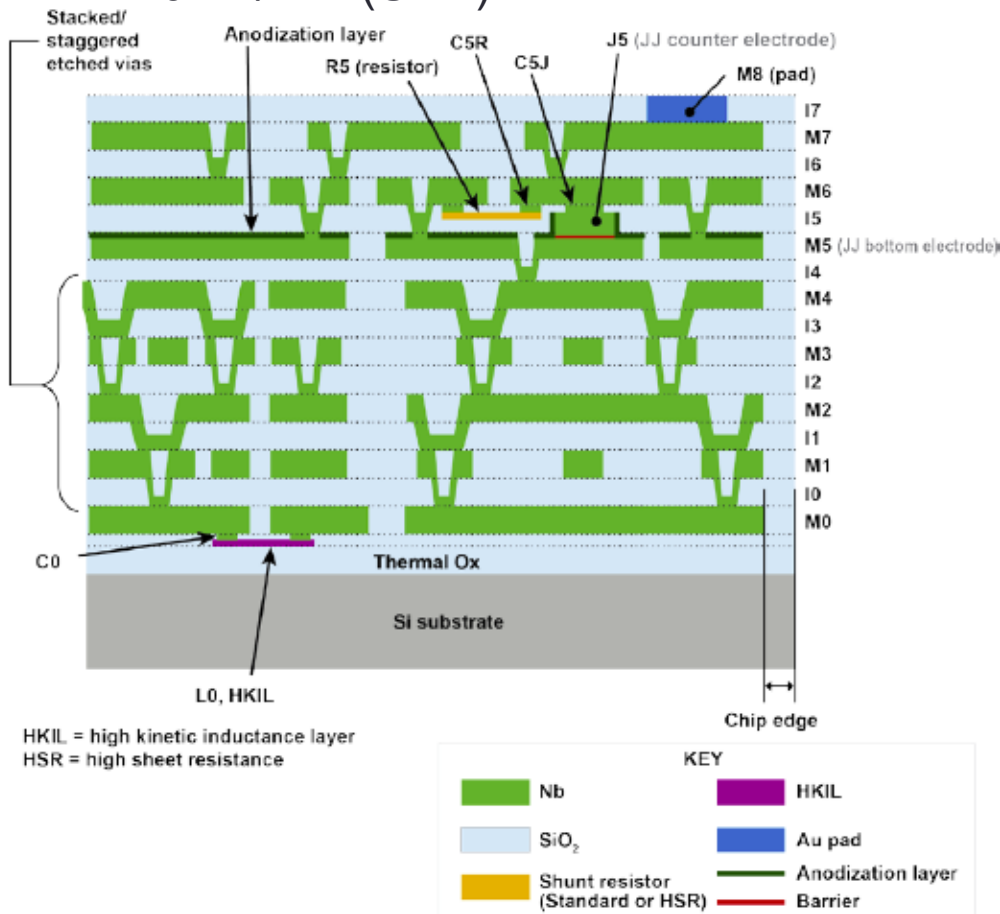
**PTLs are a highly appealing feature in RSFQ circuits.**

[5] Takagi, Katsumi, et al. "SFQ propagation properties in passive transmission lines based on a 10-Nb-layer structure." *IEEE Transactions on Applied Superconductivity* 19.3 (2009): 617-620.

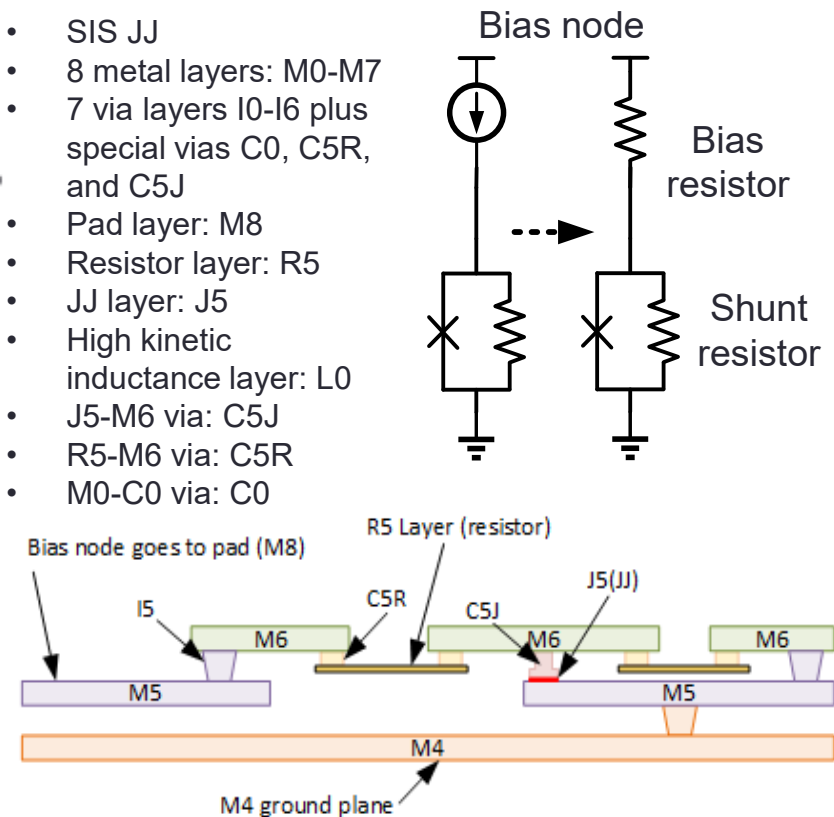
[6] P. I. Roux, K. Jackman, J. A. Delport and C. J. Fourie, "Modeling of Superconducting Passive Transmission Lines," in *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 5, pp. 1-5, Aug. 2019, Art no. 1101605.

# Fabrication Technology

- Existing fabrication processes: ISTECH-AIST, 2.5 kA/cm<sup>2</sup> (Japan), MIT-LL 10 kA/cm<sup>2</sup> (USA).



- SIS JJ
- 8 metal layers: M0-M7
- 7 via layers I0-I6 plus special vias C0, C5R, and C5J
- Pad layer: M8
- Resistor layer: R5
- JJ layer: J5
- High kinetic inductance layer: L0
- J5-M6 via: C5J
- R5-M6 via: C5R
- M0-C0 via: C0



[7] S. K. Tolpygo, V. Bolkhovsky, D. E. Oates, R. Rastogi, S. Zarr, A. L. Day, T. J. Weir, A. Wynn, and L. M. Johnson, Superconductor electronics fabrication process with MoN<sub>x</sub> kinetic inductors and self-shunted Josephson junctions, IEEE Trans. Appl. Supercond. 28, 1100212 (2018).

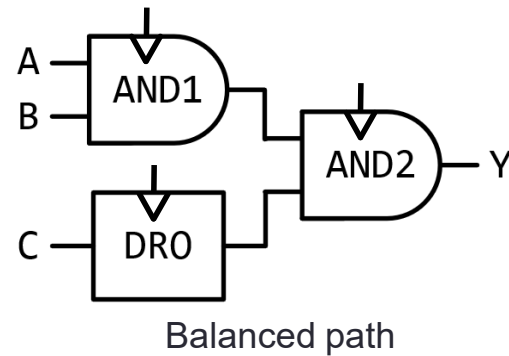
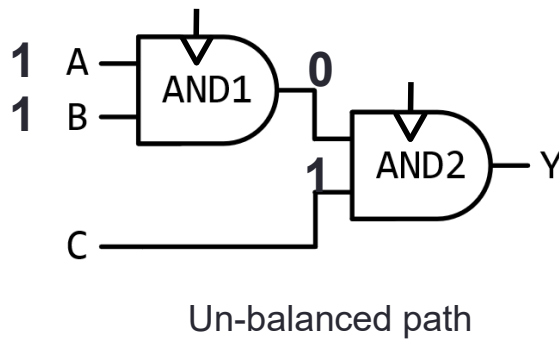
# Unique Characteristics of SFQ Logic

- Different active and passive components
  - Two-terminal JJ's and inductors vs. 3-terminal transistors and capacitors in CMOS
- Superconducting interconnect
  - Josephson transmission lines (JTLs) and passive transmission lines with small series resistors to prevent flux storage (PTLs) vs. lossy interconnect in CMOS
- Different suites of basic cells
  - Simple clocked cells (NOT, AND, OR, XOR, DFF, NDRO) and some clockless cells (delay cells, splitters) vs. complex multi-input logic cells in CMOS
- Different drive capability
  - Very low fanout count (typically just one fanout) vs. large fanout drive capability in CMOS
- Full-path balancing requirement
  - The input signals of each cell must arrive within one clock window for the correct operation of SFQ circuits
- Expensive to implement feedback loops and conditionals
- No dense on-chip memory
- Large area
  - Theoretical estimation of the maximum density of SFQ-based circuits utilizing geometric inductance of wires gives  $\sim 10^7$  JJ/cm<sup>2</sup>. Further decrease of the linewidth and spacing is problematic because of a nearly exponential growth in the mutual inductance and crosstalk between the inductors. Kinetic inductances help, but their effectiveness at smaller scales is questionable [8b].

[8b] I. I. Soloviev, V. I. Ruzhickiy, S. V. Bakurskiy, N. V. Klenov, M. Yu. Kupriyanov, A. A. Golubov, O. V. Skryabina, and V. S. Stolyarov, Superconducting digital circuits without inductors, arXiv.org > cond-mat > arXiv:2011.05856v1.

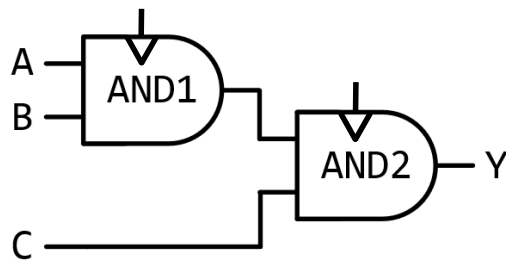
# Gate-level pipelining

- Since each gate has a clock, all the input should travel the same cycle to arrive at any gate
- The most common way to balance the path is by adding DRO cells. However, there are other ways to achieve this goal, such as optimizing the synthesis process

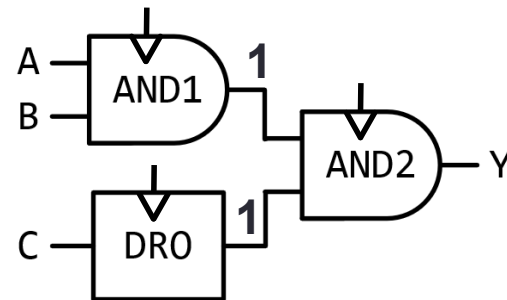


# Gate-level pipelining

- Since each gate has a clock, all the input should travel the same cycle to arrive at any gate
- The most common way to balance the path is by adding DRO cells. However, there are other ways to achieve this goal, such as optimizing the synthesis process



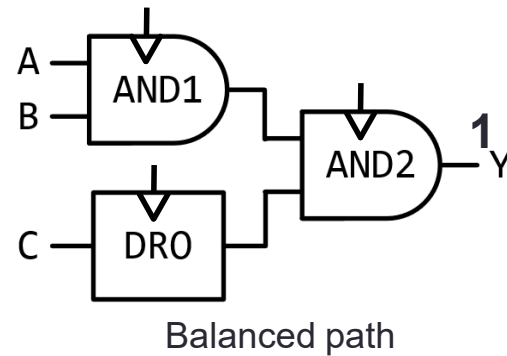
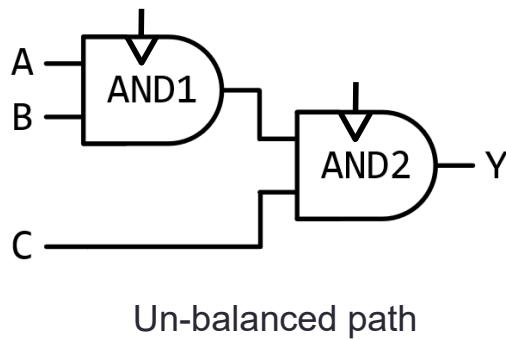
Un-balanced path



Balanced path

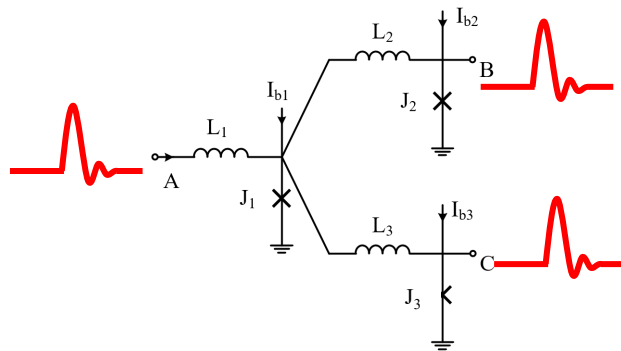
# Gate-level pipelining

- Since each gate has a clock, all the input should travel the same cycle to arrive at any gate
- The most common way to balance the path is by adding DRO cells. However, there are other ways to achieve this goal, such as optimizing the synthesis process

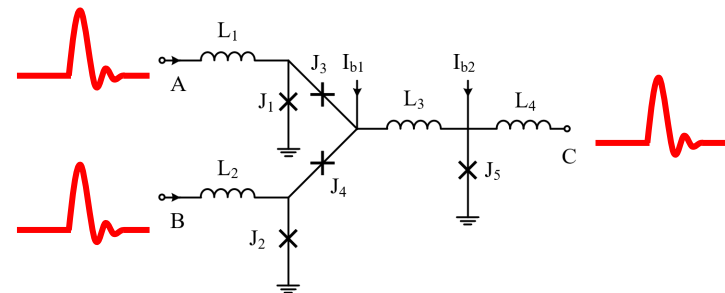


# Fanouts

- In SFQ logic, fan-out is not free. A splitter can duplicate a pulse and then drive two pins
- In SFQ logic, a merger makes two SFQ pulses drive the same pin possible



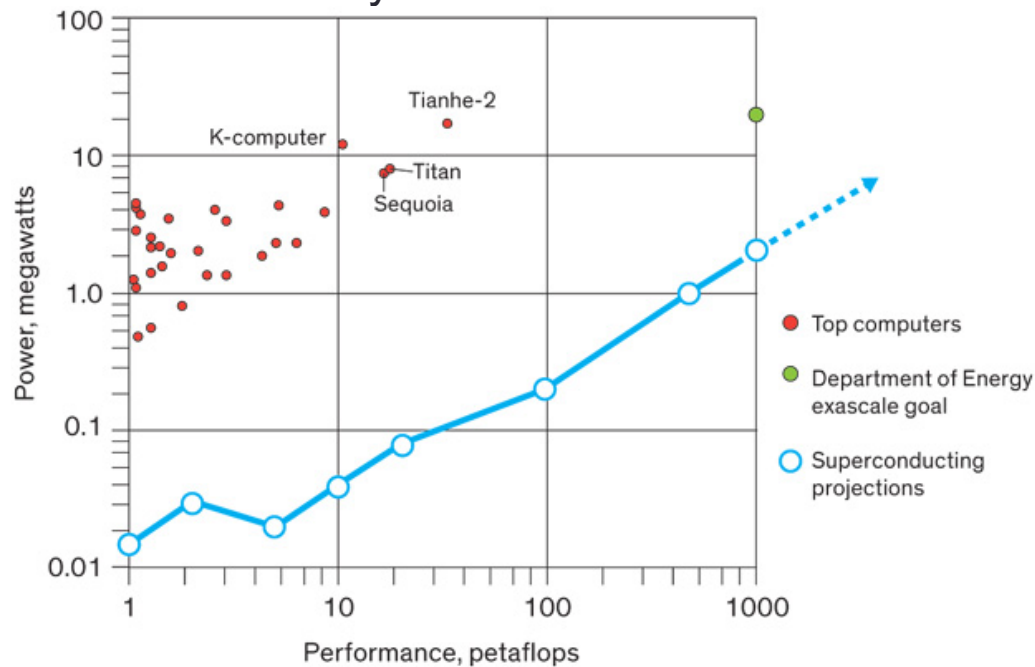
Splitter Cell



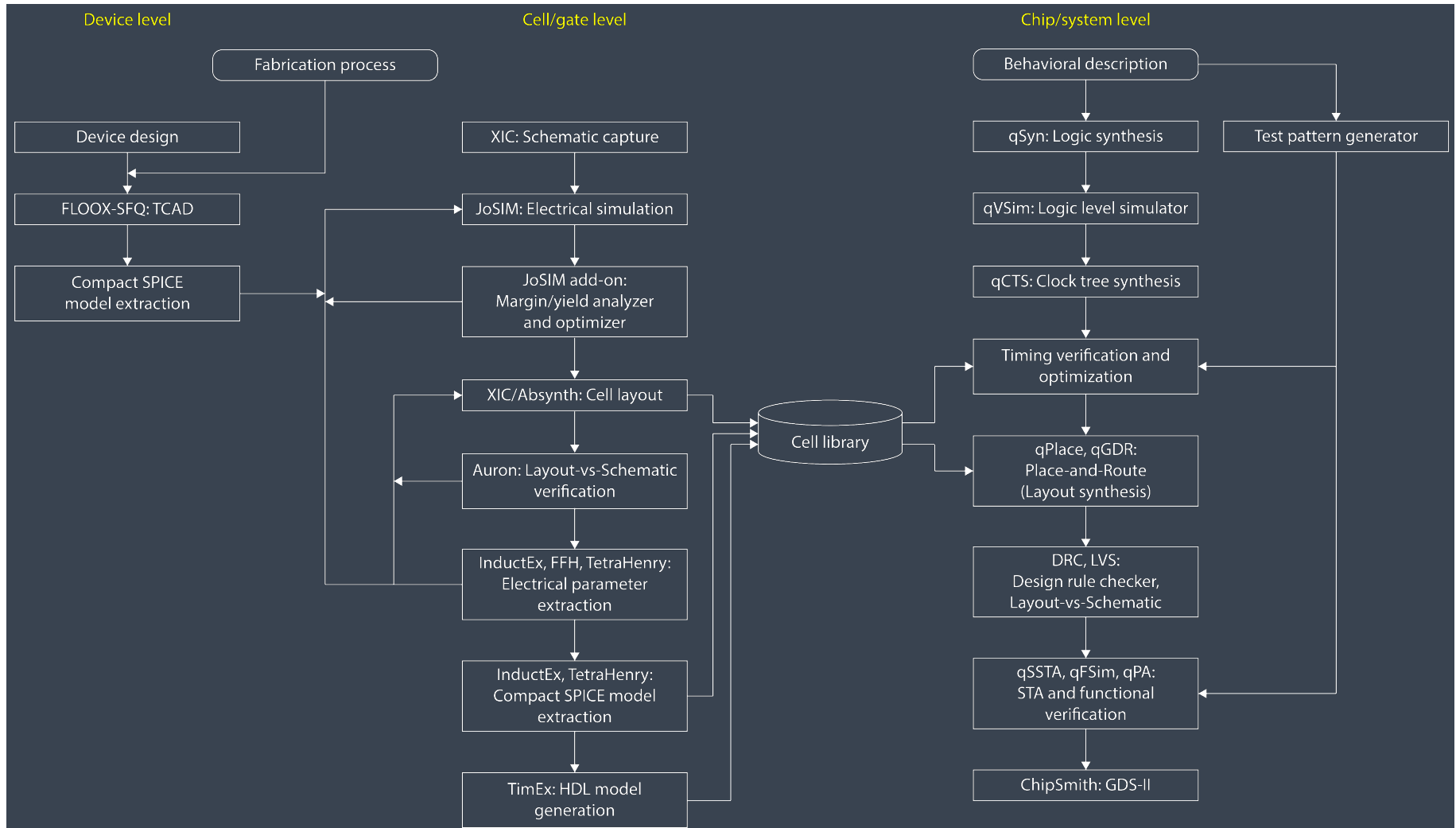
Merger Cell

# Cooling Cost

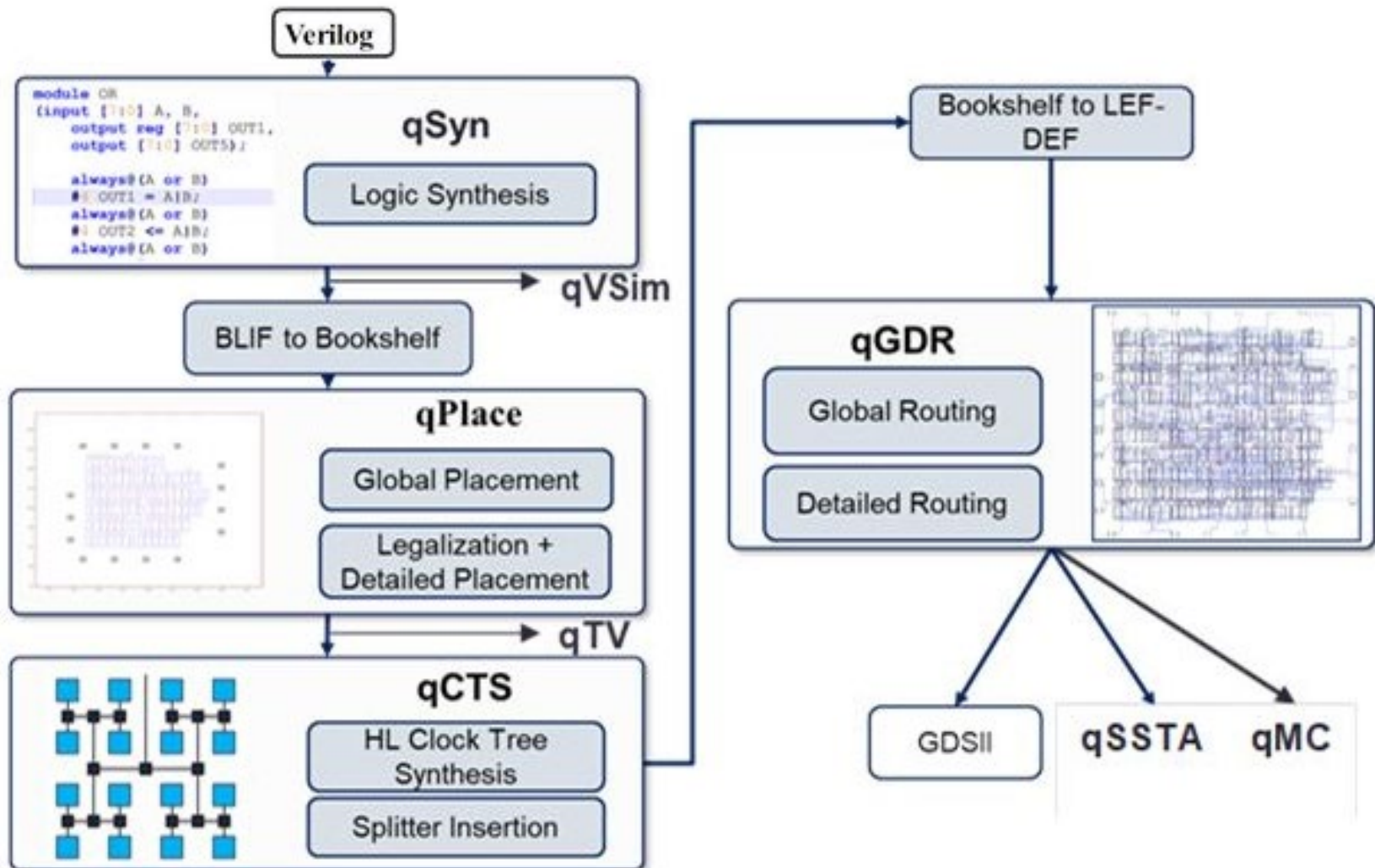
One big question has been how much the energy needed for cooling will increase a superconducting computer's energy budget. But advocates suggest it might not be much. The power drawn by commercial cryocoolers leaves "considerable room for improvement," Elie Track and Alan Kadin of the IEEE's Rebooting Computing initiative recently wrote. Even so, they say, "the power dissipated in a superconducting computer is so small that it remains 100 times more efficient than a comparable silicon computer, even after taking into account the present inefficient cryocooler."



# IARPA SuperTools Program: ColdFlux Tools

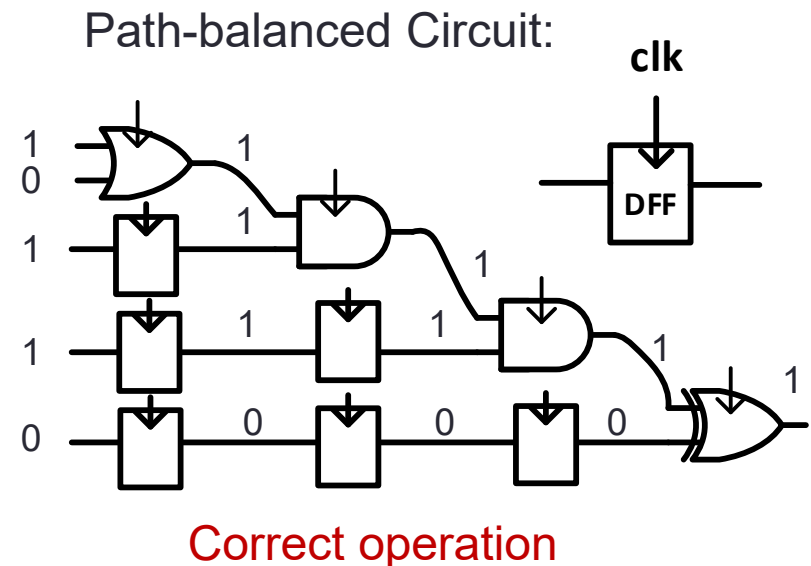
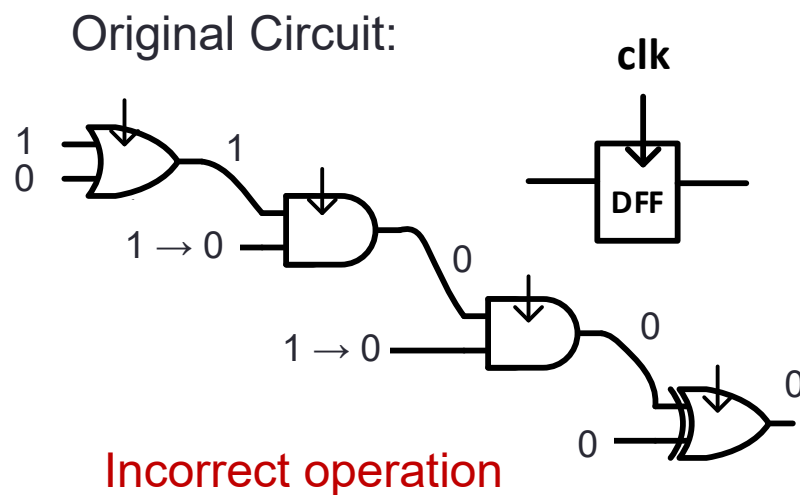


# qPALACE: ColdFlux Frontend Design Tools



# Full Path Balancing Requirement

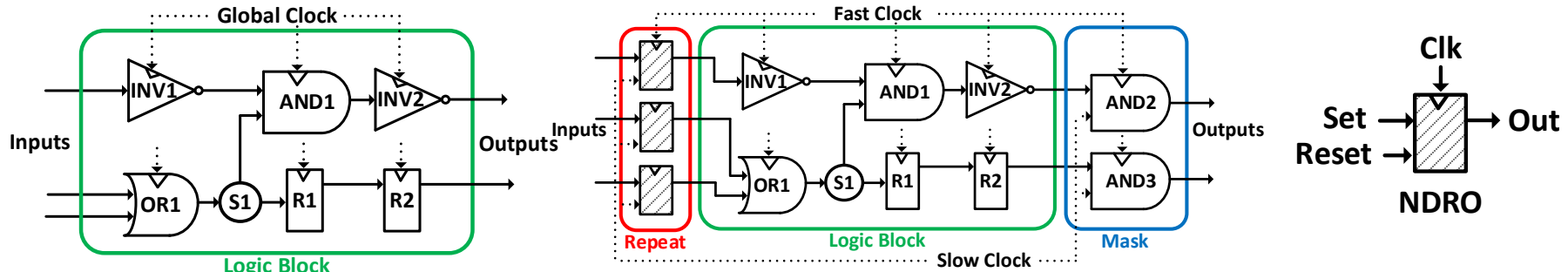
- Gate-level pipeline:
  - Most of SFQ gates (except a few gates such as splitter, merger, TFF, etc.) require receiving a clock signal to transfer the stored quantum flux to their output. This data transfer resets the superconducting current loop and destroys internal state of the logic gate. Need specialized Non-Destructive Read Out (NDRO) gates to maintain internal state after readout.
- Full Path Balancing (FPB):
  - Due to the gate-level pipelining in SFQ, length of any path from any primary input to a gate in terms of the clocked elements should be the same.





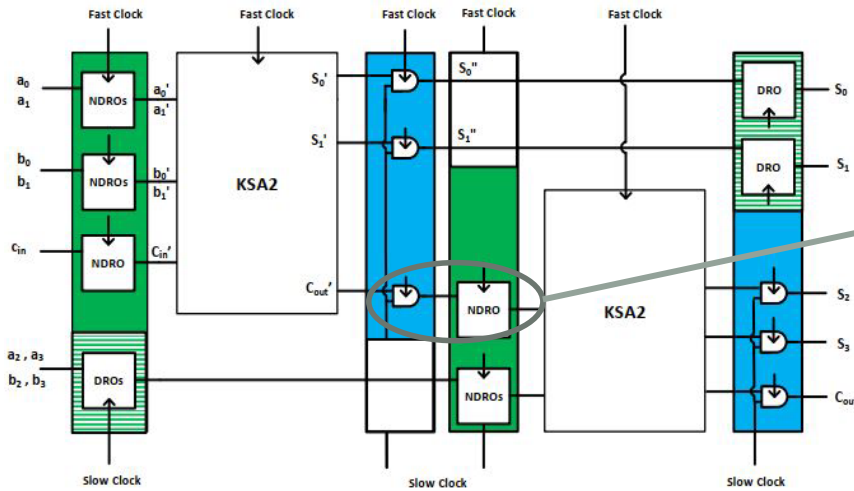


# Clocking

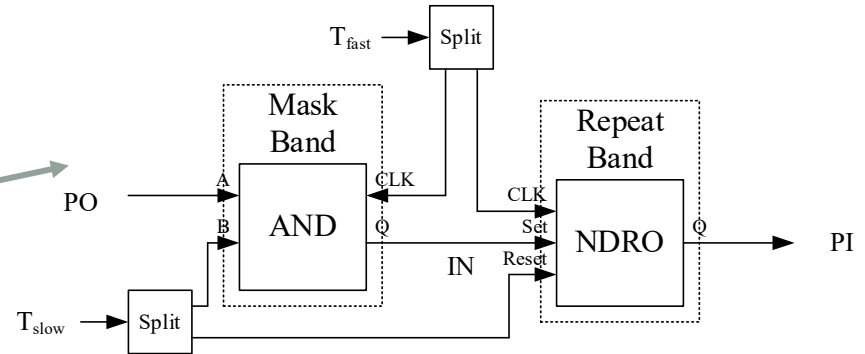


- Single clock architecture: all clocked cells are fed with the same clock to
  - The input signals for correct operation must arrive within a clock window for each cell
- Dual clock architecture [10]: clocked cells are fed with fast and slow clock signals to form a **partially path balanced (PPB)** circuit
  - Frequency relation:  $f_{fast} = (\Theta + 1) f_{slow}$  where  $\Theta \in \mathbb{Z}^+ \cup \{0\}$
  - The input signals for correct operation arrive within  $\Theta + 1$  clock windows for each cell
- Balancing register count can be reduced with the increase of  $\Theta$

# Supporting a Dual Clocking Architecture



All other paths are driven by the fast clock.



$$T_{slow} = T_{fast}(\lambda + 2)$$

Example of  $\lambda = 1$ ,  $T_{slow} = T_{fast}(\lambda + 2) = T_{fast} * 3$

## Definitions:

slack: delay of pulses between fast and slow clock sources

$LP_{slow}$ : a clock path from slow clock works as the launch path

$CP_{fast}$ : a clock path from fast clock works as the capture path

$PD_{AND}$ : propagation delay of AND

$Reset_{NDRO}$ : reset delay of NDRO

## Timing constraints:

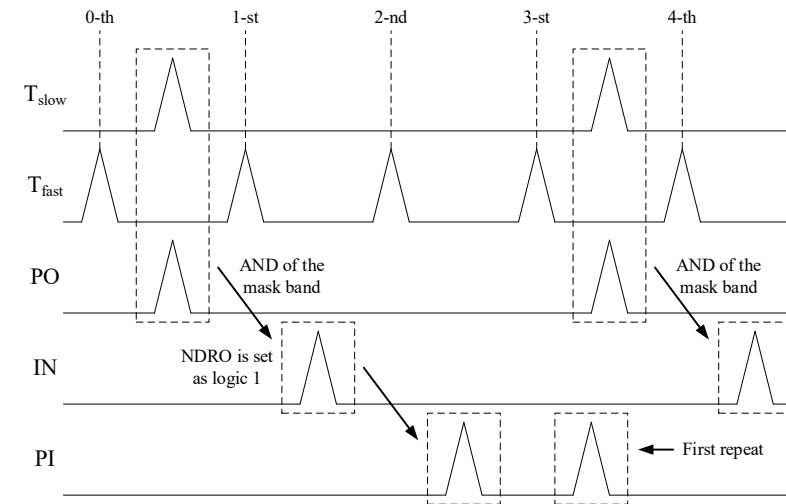
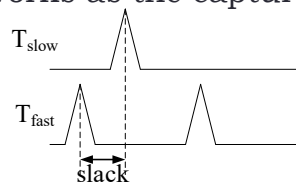
$$slack + LP_{slow} \leq CP_{fast} + T_{fast} - Setup_{AND}$$

$$slack + LP_{slow} \geq CP_{fast} + Hold_{AND}$$

$$slack + LP_{slow} \leq CP_{fast} + T_{fast} - Setup_{NDRO}$$

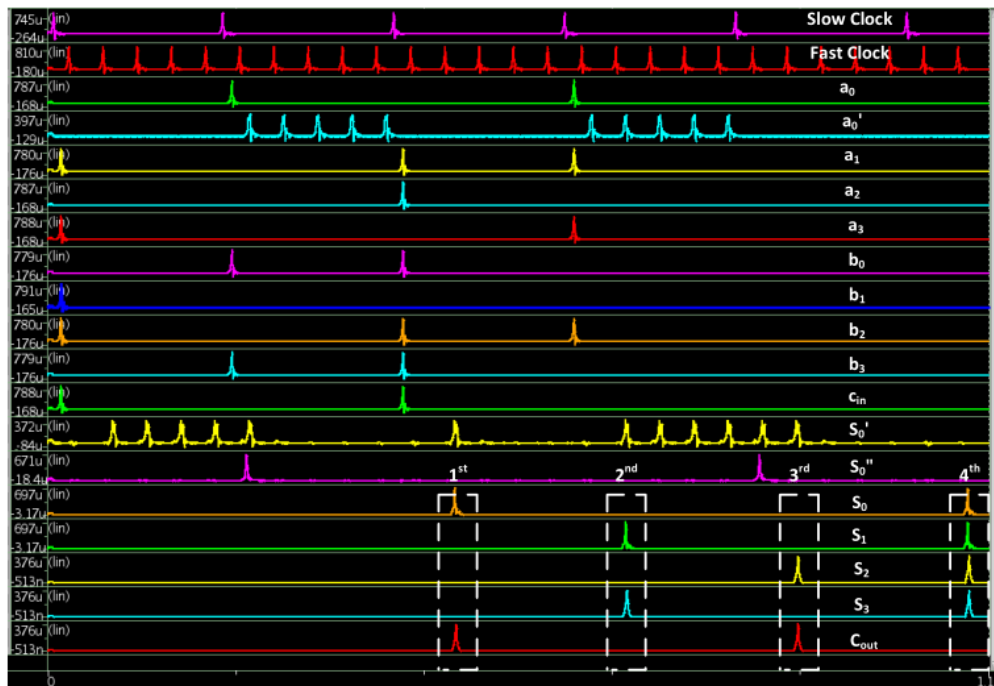
$$slack + LP_{slow} \geq CP_{fast} + Hold_{NDRO}$$

$$slack + LP_{slow} + Reset_{NDRO} \leq CP_{fast} + T_{fast} + PD_{AND}$$

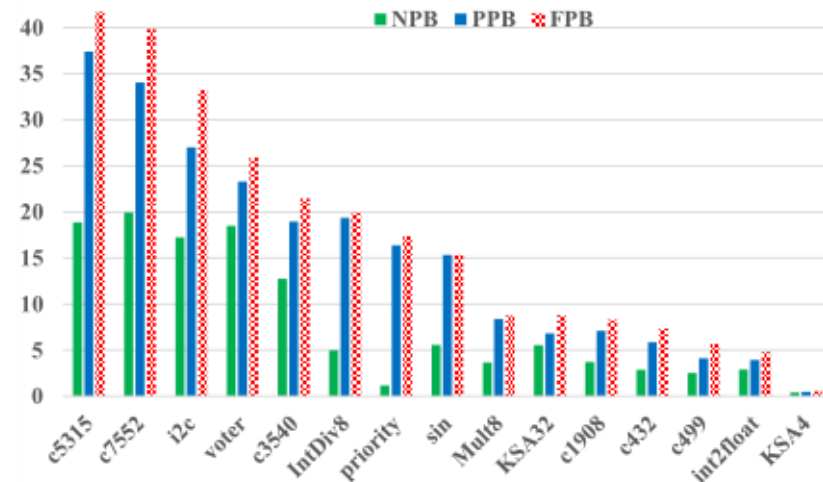


# Example of the Dual Clocking Architecture

- 4-bit adder:  $\lambda=4$ 
  - Input values:  $a_0=0101$ ,  $a_1=1011$ ,  $a_2=0010$ ,  $a_3=1001$ ,  $b_0=0110$ ,  $b_1=1000$ ,  $b_2=1011$ ,  $b_3=0110$ ,  $c_{in}=1010$ .
  - Output values:  $S_0=1001$ ,  $S_1=0101$ ,  $S_2=0011$ ,  $S_3=0101$ ,  $C_{out}=1010$ .



On average, NPB (no path balancing) and PPB (partial path balancing) reduce area by 2× and 15.38%, respectively, compared to FPB (full path balancing) case.



# Conclusion

- Under the ColdFlux project we have been developing:
  - EDA flows and tools for better design, optimization and layout of SFQ circuits in magnetic environments, with RSFQ, ERSFQ and AQFP logic circuits
  - The open-source design framework and tool suite will be made available to the public!
- In spite of much progress, key challenges exist in way of physical scaling, controlling stray electromagnetic fields, supporting multiple clock, doing current recycling, design centering to improve operating margins, designing dense on-chip memory, etc.