Title
LC Resonant Clock Resource Minimization using Compensation Capacitance

Permalink
https://escholarship.org/uc/item/3pr1r24g

Authors
Guthaus, Matthew
Lin, Ping-Yao
Fahmy, Hany
et al.

Publication Date
2015-05-01

Peer reviewed
LC Resonant Clock Resource Minimization using Compensation Capacitance

Ping-Yao Lin, Hany Ahmed Fahmy, Riadul Islam, Matthew R. Guthaus
Department of Computer Engineering, University of California, Santa Cruz
{plin11, hfahmy, rislam, mrg}@ucsc.edu

Abstract—Distributed-LC resonant clock distribution is a viable technique to reduce clock distribution network (CDN) dynamic power. However, resonant clocks can require significant on-chip resources to form the inductors and decoupling capacitors which discourages adoption. This paper uses a compensation capacitor ($C_c$) to reduce the overhead of the on-chip inductor and capacitor resources while still maintaining 49.9% power over traditional buffered clocks.

I. INTRODUCTION

The clock distribution network (CDN) can consume 30-70% of the entire chip power due to the high switching rate and large capacitance. One way of decreasing this dynamic power is implementing resonant clocking [1]. Distributed-LC resonant clocking circumvents this dynamic power barrier by oscillating energy between an electrical potential and magnetic field in a parallel capacitance ($C_{clk}$) and inductance ($L$), respectively. An additional decoupling capacitance ($C_d$), serving as AC ground, offsets the voltage range to 0−$V_{dd}$ by forming a series LC tank as shown in Figure 1(a).

Industry has demonstrated power savings of these parallel LC-tank clocks [2] and confirms stable phases and magnitudes compared to standing wave [3] and rotary/salphasic [4] resonant clocks. Academic algorithms have further optimized these resonant clocks [5] and have shown to save up to 90% of the dynamic power at the expense of large inductor and decoupling capacitor on-chip area [6]. The inductor area is perhaps the most significant obstacle for significant power savings [5], [6]. In addition, the decoupling capacitance is an extra cost that is often neglected. This paper proposes a modified LC-tank topology that uses a compensation capacitor ($C_c$) to decrease the total inductor and capacitor area while still maintaining resonant power savings. Specifically, this paper is the first to:

- Minimize inductor and capacitor overhead in LC resonant clocks using a coupling capacitor.
- Derive the theoretical formulations how $C_c$ and $C_d$ affect resonant frequency.
- Provide an approximate analysis using the Miller model of $C_c$ on resonant frequency.

The remainder of this paper proceeds as follows: Section II presents necessary background on on-chip inductance/capacitance and resonant theory. Section III introduces the theoretical formulation behind coupling capacitance in resonant clocks.

II. BACKGROUND

A. Integrated Inductors

On-chip spiral inductors can be easily manufactured in standard CMOS processes. Also, high-Q spiral inductors are available with modern processes that have thick (~10um) oxides [7]. However, on-chip inductors come with parasitic resistance and capacitance that decrease their quality factor. The parasitic resistance can be included in the model of the inductor, and the parasitic capacitance can be treated as part of the decoupling capacitance in a resonant clock [5]. The inductance of a square spiral inductor, as shown in Figure 1(b), can be approximated as

$$L = 0.002l[\ln \frac{2l}{w + t} + 0.5 + \frac{w + t}{3l}], \quad (1)$$

where $l$ is the length of trace, $w$ is the width of trace, and $t$ is the thickness of the metal [8]. Given the physical parameters, the area of the inductor is

$$Area = d_o^2 - (d_i + 2n(s + w))^2, \quad (2)$$

where $d_o$ is the outer diameter, $d_i$ is the inner diameter, $n$ is the number of turns, and $s$ is the space between turns [8].

B. Integrated Capacitors

Multiple types of on-chip capacitors are utilized in ICs, such as MOS (metal-oxide-semiconductor), PIP (polysilicon-insulator-polysilicon), MIM (metal-insulator-metal), and lateral flux capacitors. These capacitors have different design and
performance characteristics. Historically, MIM capacitors have been used in RF and mixed-signal ICs due to low leakage, high linearity, low process variations, and low temperature variations. MIM capacitors with a capacitance density comparable to MOS capacitors have been achieved by reducing the dielectric thickness and employing high-k dielectrics.

The benefit of MOS capacitors is the compatibility with CMOS technology and a high capacitance density. However, MOS capacitors exhibit challenges due to non-linearity, strong voltage dependence, and high leakage. Overall, the capacitive density of $10–20 fF/µm^2$ can be fabricated in modern CMOS technologies [9]. Much like inductors, the parasitic resistance in capacitors influences the efficiency of a decoupling capacitor.

C. Resonant Theory

These on-chip inductors and capacitors can be used to form a series or parallel LC “tank” circuit which resonates at a particular frequency. This so-called resonant frequency of an ideal LC tank is the frequency when the network has zero total reactance,

$$ f_0 = \frac{1}{2\pi\sqrt{LC}}, \quad (3) $$

where $L$ is the inductance and $C$ is either the $C_{clk}$ or $C_d$ in our case. When the impedance of an LC tank is near infinite due to the reactances cancelling, the drive capability requirement of the clock buffer is minimized and clock energy can be saved.

As shown in Figure 1(a), our resonant clock topology contains two LC tanks: one parallel and one series. The parallel LC tank will provide infinite impedance at resonance while the series impedance will provide zero impedance. These can counteract each other if they resonate at the same frequency. Therefore, $C_d$ is typically much larger than $C_{clk}$ to separate the resonant frequencies of the parallel $LC_{clk}$ and extra series $LC_d$ tanks according to

$$ \frac{1}{2\pi\sqrt{LC_d}} << \frac{1}{2\pi\sqrt{L'C_{clk}}} \quad (4) $$

This ensures a wide margin such as $f_{clk}/f_d \sim 3$ by making $C_d$ usually $10 \times C_{clk}$ or more.

III. Compensation Capacitance

As illustrated in Figure 2, compensation capacitance ($C_c$) could be used in two possible topologies: one position is in parallel with $C_{clk}$ while the second is in parallel with $L$. If $C_c$ is in parallel with $C_{clk}$, the size of inductance can be reduced by adding to the capacitance in Equation (3) while retaining the same resonant frequency ($f_0$). On the other hand, if $C_c$ is chosen in parallel with $L$, the Miller Effect [10] suggests that $C_c$ can be divided as $C_{c1}$ and $C_{c2}$ in parallel and effectively added to $C_{clk}$ and $C_d$. This would, in turn, reduce the necessary sizes of both $L$ and $C_d$.

If both the parallel and series tank circuits are considered in Figure 1(a), the resonant frequency from Equation 3 is more accurately given as

$$ f_{clk} = \frac{1}{2\pi} \sqrt{\frac{C_{clk} + C_d}{LC_{clk}C_d}}. \quad (5) $$

If $C_d = 10 \times C_{clk}$, this reduces to

$$ f_{clk} = \sqrt{\frac{1}{10}} \cdot \frac{1}{2\pi\sqrt{LC_{clk}}}, \quad (6) $$

which shows a $\sqrt{10}/10 = 4.9\%$ increase in resonant frequency due to the parasitic series $LC_d$ tank. This can be utilized to reduce inductor size and save area at a fixed frequency.

If $C_c$ is added in parallel with $C_{clk}$, the resonant frequency will add $C_c$ to $C_{clk}$,

$$ f'_{clk} = \frac{1}{2\pi} \sqrt{\frac{C_{clk} + C_c + C_d}{L(C_{clk} + C_c)C_d}}. \quad (7) $$

If compared to Equation (5), this provides a decrease in resonant frequency.

If we consider adding $C_c$ in parallel with $L$, the resonant frequency will add $C_c$ to $C_{clk}$ and $C_{c2}$ to $C_d$,

$$ f'_{clk} = \frac{1}{2\pi} \sqrt{\frac{(C_{clk} + C_{c1}) + (C_d + C_{c2})}{L(C_{clk} + C_{c1})(C_d + C_{c2})}}, \quad (9) $$

where $C_{c1} = Cc(1 - A_v)$, $C_{c2} = C_c(1 - A_v^{-1})$, and $A_v$ is the ratio of voltage after the $L$ to before the $L$. Here, $A_v$ is an approximation to calculate $f'_{clk}$. On the other hand, applying $C_c$ directly to the formulation, the resonant frequency is

$$ f'_{clk} = \frac{1}{2\pi} \sqrt{\frac{C_{clk} + C_d}{L(C_{clk}C_d + C_c(C_{clk} + C_d))}}, \quad (10) $$

which gives a decrease in resonant frequency. Since Equation (10) is strictly less than Equation (7), this provides more savings at all times.

A significant difference in the topologies is an additional capacitive effect ($a_2$) of $C_{c2}$ to $C_d$ when $C_c$ is in parallel with $L$, where $a_2 = (f_d/f_0)^2$. In this case, applying $C_c$ to the formulation, the decoupling resonant frequency is

$$ f'_{d} = \frac{1}{2\pi\sqrt{L(C_d + C_c)}}, \quad (12) $$

which is a factor $\sqrt{a_2}$ lower than the case without $C_c$. Consequently, it offers the advantage of sharing $C_c$ directly from $C_d$ for the same $f_d$. 

Fig. 2. Possible $C_c$ positions to adjust resonant frequency in resonant clocks.
Based on the resonant frequency in Equation (3), \( L \) can be reduced by adding \( C_c \) and considering the Miller approximation to get \( C_{c1} \) and \( C_{c2} \). If \( C_{c1} + C_{clk} = a_1 \cdot C_{clk} \), where \( a_1 \) is the ratio between capacitances, then \( L \) can be reduced as \( L/a_1 \) for the same resonant frequency \( (f_{clk}) \).

\[
L' = \frac{L}{a_1} = \frac{L}{(f_{clk}/f'_{clk})^2}, (13)
\]

where \( a_1 = (f_{clk}/f'_{clk})^2 \) is the capacitive effect from \( C_{c1} \), and \( L' \) and \( f'_{clk} \) are the inductance and resonant frequency after adding \( C_c \), respectively.

Taking the derivative of Equation (13) for \( C_c \) in parallel with \( C_{clk} \), we get

\[
\frac{dL'}{dC_c} = L \frac{C_{c1} + C'_{c} - 1}{(1 + \frac{C_{c1}}{C_{clk}})^2}, (14)
\]

which is a negative slope with sharp or smooth \( L \) curve as \( C_c \) value decreases or increases. Doing the same derivative for \( C_c \) in parallel with \( L \), we get

\[
\frac{dL'}{dC_c} = L \frac{1 - \frac{1}{C_{c1}} - \frac{1}{C_{clk}}}{(1 + \frac{C_{c1}}{C_{clk}} + \frac{C_{c1}}{C_{c1}})^2}, (15)
\]

which is a negative, steeper slope for small \( C_c \) values. Therefore, \( C_c \) in parallel with \( L \) is a more effective topology in terms of reducing \( L \).

IV. EXPERIMENTS

The experimental results are divided into two parts. All results are measured using HSPICE. The first part uses a buffered monolithic LC-tank to verify our theoretical analysis in Section III. The capacitance and inductance values are selected to represent a lumped CDN. The baseline resonant frequency \( (f_0) \) and \( V_{dd} \) are set to be 2GHz and 1V, respectively. The second part utilizes the concept of \( C_c \) insertion on a distributed resonant clock with the industrial-representative ISPD 2010 benchmarks from IBM and Intel [11]. The ISPD benchmarks are in a 45nm technology and use \( V_{dd} \) of 1V and \( f_0 \) of 1GHz. The resonant benchmark designs are synthesized using the Resonant Clock Synthesis (ROCKS) methodology [6] and the C++ source code is modified for our new circuit techniques.

A. Monolithic LC-tank

The first experiment verifies our previous theories using a monolithic LC-tank resonant clock with \( C_c \) insertion like Figure 2. A single LC tank loads a lumped clock grid \( (C_{clk} = 5pF) \), while \( L = 1.39nH \) to achieve a desired resonant frequency of 2GHz. \( C_c \) was chosen to be 0.5\( C_{clk} = 2.5pF \) for demonstration, but as shown in Equation (14)-(15) the trends hold over all sizes.

The frequency analysis results are summarized in Table I. These verify that \( C_c \) in parallel with \( L \) enables a lower resonant frequency than in parallel with \( C_{clk} \) due to the higher capacitive effects \( (a_1 \) and \( a_2) \). Explicitly, \( f_{clk} \) with \( C_c \) in parallel with \( L \) shows 3.8% lower frequency than in parallel with \( C_{clk} \), same as the theoretical models suggested in Section III. The parasitic decap resonant frequency \( (f_d) \) with \( C_c \) in parallel with \( L \) also matches the theoretical results in Equation (12).

![Fig. 3. A single LC tank loads a lumped clock grid (LC). The resonant frequency in Equation (3) results in a slightly tighter \( f_{clk}/f_d \) ratio.](image)

**Fig. 3.** For a 2GHz clock, \( C_c \) inserted in parallel with \( L \) saves more in terms of \( L \) and on-chip area, Area\((L) + \) Area\((C_c) + \) Area\((C_d)\).

Instead of changing resonant frequency, \( L \) and/or \( C_d \) can be reduced while maintaining the same frequency. For \( C_c \) in parallel with \( L \), the capacitive effects are higher and \( C_d \) can be reduced in addition to \( L \). Specifically, the capacitive effect \( (a_1) \) from \( C_{c1} \) suggests that \( L \) can be reduced as \( L/a_1 = 1.39/1.55 = 0.90nH \) which also increases \( f_d \). The trade-off between the \( L \) reduction and the range of resonant frequency separation \( (f_d vs f_{clk}) \) is based on Equation (3) resulting in a slightly tighter \( f_{clk}/f_d \) ratio. On the other hand, \( C_d \) could alternatively be reduced by 2.5pF which is equal to the inserted \( C_c \) and results in extra resource sharing. However, this is at the expense of higher \( f_d \).

Finally, when we consider the actual on-chip inductor and capacitor areas as in Section II, Figure 3 shows the relationship between various \( C_c \) values and the total area \( (Area(L) + Area(C_c) + Area(C_d)) \) along with the power consumption. We assume that our inductor has \( w = 15\mu m, s = 15\mu m, \) and \( d_i = 100\mu m \). We also assume a capacitive density of \( 10fF/\mu m^2 \) for a typical MOS capacitor.

As seen in Figure 3, \( C_c \) insertion in parallel with \( L \) saves more in terms of \( L \) and on-chip area for all cases. The downside of this is that the power consumption goes up in both cases, but is explicitly lower when inserting \( C_c \) in parallel with \( L \) in all cases. This is because the voltage swing of the decap \( C_d \) is miniscule whereas \( C_{clk} \) has rail-to-rail swing. In our example \( C_c = 2.5pF, 21.4% \) total area is saved with only a 2.1% power increase and 10.4% inductor Q decrease when \( C_c \) is parallel with \( L \). If we consider efficiency of saving area, the knee of the curve is at \( C_c = 10pF \) which offers area savings of 44.7% with a power increase of 5.6% and inductor Q decrease of 29.5%.

B. Distributed LC-tanks

We then utilize \( C_c \) insertion in parallel with \( L \) in a distributed resonant clock synthesis (ROCKS) [6] methodology. ROCKS starts similar to Meshworks [12] by creating an initial buffered grid. ROCKS then places and sizes LC tanks to form a resonant grid using an iterative LC tank clustering algorithm. After this, ROCKS performs a frequency-based grid buffer
sizing. Finally, a top-level buffered tree is generated with minimum wire length DME [13] to drive the resonant grid.

The $C_c$ insertion is targeting the knee of curve in Figure 3 to save the most area. Table II organizes the power and inductor area analysis results and compared with a non-resonant grid, ROCKS, and our results with $C_c$ insertion. For accurate comparison with ROCKS, the intrinsic resistances of inductor and capacitor are not taken into account since parasitic resistances of clock tree and grid dominate.

The inductor Q can be found by using $Q = \omega L / R_s$, where $R_s$ is the parasitic resistance in series with $L$. $R_s$ is obtained from ASITIC [8], assuming a single-layered square inductor model with no ground plane. A large Q can store more energy while a low Q does the opposite. Our results demonstrate that a distributed $C_c$-enhanced resonant clock reduces the inductor area overhead by approximately 11.8%, decreases the inductor Q about 19.6%, and still maintains nearly 49.9% power savings. This indicates that a $C_c$-enhanced LC circuit keeps the high amplitude at resonance, with a wider bandwidth around the resonant frequency.

V. Conclusions

Resonant clocks using distributed LC tanks hold great promise for power reduction by performing on-chip energy recycling in the form of the electrical and magnetic fields that resonate between the clock capacitance and an on-chip inductor. However, the size and area of inductance ($L$) and decoupling capacitance ($C_d$) are an extra cost that many designers wish to avoid. This paper demonstrated the concept of a compensation capacitor ($C_c$) to reduce this overhead in a resonant clock circuit. Our results show that $L$ can be reduced most effectively by placing $C_c$ in parallel with $L$ rather than as additional clock capacitance load. This has the advantage of higher capacitive effects which result in both smaller $L$ and $C_d$.

<p>| TABLE I. $C_c$ insertion in parallel with the inductor enables a lower resonant frequency due to the higher capacitive effects ($a_1$ and $a_2$). |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>$f_{res}$ (GHz)</th>
<th>$C_c$ (nF)</th>
<th>$Q$</th>
<th>$f_{res}$ (GHz)</th>
<th>$C_c$ (nF)</th>
<th>$Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>1.39</td>
<td>603.96</td>
<td>1.00</td>
<td>1.39</td>
</tr>
<tr>
<td>1.00</td>
<td>1.39</td>
<td>0.99</td>
<td>301.75</td>
<td>1.00</td>
<td>0.99</td>
</tr>
</tbody>
</table>

| TABLE II. $C_c$ implementation with distributed LC-tanks on ISPD 2010 benchmarks shows overhead reduction and maintains high power saving. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Non-resonant CDN | ROCKS [6] | $C_c$ implementation |
| Sink | Cap | Chip Area | Pwr\* | Pwr\* | Pwr\* | LA/CA | Pwr\* | Pwr\* | LA/CA | Q | Pwr\* |
|-------|------|-------------|------|------|------|-------|------|------|-------|------|-------|------|
| # | pF | mm² | mW | mW | ps | % | % | % | mW | ps | % | % |
| 01 | 1107 | 18.9 | 64.0 | 327.2 | 203.7 | 41 | 29.2 | 37.7 | 204.2 | 41 | 17.1 | 18.6 | 37.6 |
| 02 | 2249 | 39.2 | 91.0 | 600.4 | 345.3 | 36 | 27.5 | 42.5 | 356.3 | 36 | 15.9 | 18.6 | 42.3 |
| 03 | 1200 | 18.1 | 1.4 | 141.7 | 9.6 | 33 | 30.7 | 71.6 | 47.0 | 42 | 19.3 | 20.4 | 59.0 |
| 04 | 1845 | 12.3 | 5.7 | 177.2 | 45.9 | 8 | 30.0 | 52.1 | 87.4 | 9 | 19.1 | 18.7 | 48.1 |
| 05 | 1016 | 5.2 | 5.8 | 111.4 | 40.3 | 9 | 29.6 | 60.5 | 52.9 | 13 | 17.8 | 19.1 | 52.5 |
| 06 | 981 | 12.6 | 1.7 | 103.0 | 24 | 25 | 36.2 | 13.4 | 39.7 | 30 | 21.9 | 19.0 | 61.5 |
| 07 | 1915 | 18.1 | 3.5 | 168.1 | 78.3 | 9 | 31.0 | 53.4 | 82.4 | 8 | 19.2 | 19.0 | 51.0 |
| 08 | 1134 | 13.0 | 2.6 | 117.7 | 49.0 | 13 | 28.2 | 58.3 | 59.6 | 12 | 17.7 | 23.2 | 49.4 |

Average | 1431 | 17.2 | 21.9 | 213.8 | 107.7 | 22 | 30.3 | 56.2 | 114.9 | 24 | 18.5 | 19.6 | 49.9 |

Pwr\*: Switched capacitance CDN power. Pwr\* : ROCKS power in HSPICE. Pwr\* : $C_c$ implementation power in HSPICE. LA/CA: Total inductor area normalized to metal layer (chip) area. LA/CA: Reduced total inductor area normalized to metal layer (chip) area.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant CCF-1053838.

REFERENCES