Meta-Heuristic Optimization of Transistor Sizing in CMOS Digital Designs

Prashanth H C\textsuperscript{a} and Madhav Rao\textsuperscript{b}

\textit{International Institute of Information Technology Bangalore, India}

Keywords: Circuit Optimization, EDA, Single Objective Functions, Multi-Objective Functions, VLSI, Adder Design.

Abstract: Designing new custom standard cells or digital circuits using automated optimization is challenging considering the large design space, performance trade-offs and continuous technology progression. Besides, a comprehensive study and analysis of different algorithms applied towards optimizing higher-order custom digital circuit design is imperative. In this work, 28 Transistor (28T) 1-bit full-adder (FA) is designed and investigated for six optimization algorithms, including particle-swarm-optimization (PSO), evolutionary strategy (ES), genetic algorithm (GA), differential evolution (DE), NSGA-II, and NSGA-III. The algorithms are evaluated and benchmarked, considering diversity of candidate solutions, monotonicity of fitness convergence, and capability to reach the best solution when initiated with a randomly seeded solution. This work establishes that GA produces best-fit circuits among all the single-objective algorithms. ES and GA exhibit good design-space exploration, unlike PSO and DE, which are influenced by local optima. NSGA-II, and NSGA-III are preferred when the objective is to give equal importance to the targeted parameters. The extensive evaluation of the algorithms in this work will aid in adopting an effective strategy for optimizing custom circuits for the specified objective parameters.

1 INTRODUCTION

Integrated Circuit (IC) design involves many tasks to manufacture a fool-proof tape-out chip (Dey et al., 2021; Innocenti et al., 2015), of which one of the laborious processes is the transistor level optimized standard cell design (Hong et al., 2022; Eriksson et al., 2003; Lim et al., 2017; Jo et al., 2019). The transistor level design as a spice-netlist should undergo optimization to yield the best results towards the targeted parameters (Kashfi et al., 2011). Post optimization, extensive characterization of the design is performed to introduce the design as a cell in the library for system-level synthesis (Abazyan et al., 2021; Abazyan, 2021; Rahman et al., 2013; Matos et al., 2019; Cao et al., 2021b; Cao et al., 2021a). System-level synthesis is based on picking the most appropriate standard cells within the library (Matos et al., 2019). Optimizations at the synthesis stage are supposed to meet the system design requirement constraints. Currently, standard cells are limited to basic gates and have not progressed towards achieving custom cell design, which is expected to achieve better power-performance-analysis (PPA) and render compact designs (H C and Rao, 2022). The traditional approach to characterizing cell design is by configuring a parameterized spice simulation by sweeping all the parameters for the cell design to identify best solutions (Mezzomo et al., 2011; Cao et al., 2018). Most parameterized runs are automated, but these methods to trace all output parameters covering all possible input combinations become a tedious task for higher order input functions. Besides, the computing capacity required to generate and establish results through this approach is high. Hence the method is not scalable along higher-order circuits. Circuit optimization for any new higher-order custom standard cells and any custom design outside the ASIC flow is expected to establish the best hardware parameters when put on the silicon chip.

Generally, particle-swarm-optimization (PSO) (Zhan et al., 2009), and other extended forms of exact algorithms (Nikoubin et al., 2010) are popularly employed to achieve the optimal results targeted toward specific objective goals. Other algorithms in the form of evolutionary, genetic approach, NSGA (Ishibuchi et al., 2016; Deb and
Jain, 2014a), advance NSGA (Deb and Jain, 2014b), and NSGA combined with PSO (Sasikumar and Muthaiah, 2017) are not comprehensively evaluated under the same conditions for transistor sizing, although these methods are known for optimization in different set of fields (Conesa-Muñoz et al., 2012; Gu et al., 2020; Pang et al., 2020; Xiaqing et al., 2019; Zheng et al., 2016). Additionally, the optimization runs have always been studied for a single objective parameter which does not offer any insight into the other hardware metrics. The traces of the solutions in the design space only suggests a uni-dimensional approach of minimizing the targeted goal without considering its impact on the other crucial design parameters. The contradicting hardware metrics, especially performance and power or area cost, enables to settle at power-performance or footprint-performance tradeoffs inevitably.

In the past, considering the product of two objectives as a single objective to optimize the design was also attempted. However, the dominant objective parameter dictates the optimization path. Figure 1 represents a typical scenario with a set of solutions for a hardware design, and its pareto-frontier which aides in selecting the solutions. The pareto-front is generally applied for multi-objective optimization of designs targeted for conflicting objectives. Generally, candidate solutions falling close to the pareto-front are selected based on the specified metrics. Given equal importance to both objectives, such as power and delay metrics, Power and performance efficient designs are achieved by picking a pareto-optimal candidate referred to as Good Power Efficiency point as highlighted in the Figure. Similarly, two other pareto-optimal candidates satisfying the corresponding pair of objectives are annotated as Good Power Density, and Good Area Efficiency is shown in the Figure. Generally, Power and Area metrics follow a similar trend; hence, the pareto-front solution will be a point close to the origin instead of a line, as discussed later. In the past, there have been attempts to showcase a specific algorithm better than the others for circuit optimization. However, without looking at the configuration parameters required to run the exercise (Sasikumar and Muthaiah, 2017). The runtime and compute, memory required to run different algorithms to evolve a similar solution for the same hardware goals are different. Hence a particular need to evaluate different algorithms under similar configurations for the same design is desirable. As per the authors’ knowledge, the evaluation of a wide-variety of meta-heuristic algorithms for circuit optimization with the same compute resource and runtime allocation was not performed before. In this work, single and multi-objective algorithms with different objective functions are executed independently.

As shown in Figure 2, a flow was adopted to investigate the algorithm performance towards optimizing a custom circuit for hardware metrics defined as objectives. The circuit topology under investigation is fed to the flow with the desired algorithm in the form of an objective function. For all the algorithmic runs, a random solution was used as seed circuit. Circuit measurements are post-layout spice simulations. This flow remains independent of the technology progression, provided PDK files are made available for the simulation of the circuit. Based on the fitness of the candidate solutions, in terms of the objective function, the candidates are updated based on the chosen algorithm for the next generation. The new candidate designs are again measured and evaluated, and the same process continues till a termination criteria is met. Four different single-objective algorithms - Genetic algorithm (GA), Evolutionary strategy (ES), Particle swarm optimization (PSO), and Differential evolution (DE) were considered.

**Evolutionary Strategy (ES)** is a popular search algorithm incorporating selection and mutation opera-
tors in the defined computational space. A standard version with a Gaussian mutation is employed for achieving real-valued optimization, and the selection is purely based on the fitness score. Genetic Algorithm (GA) is a class of EA which represents a basic \( \mu + \lambda \) approach for single-objective problems, where \( \mu \) and \( \lambda \) represents parents, and offsprings at each generation respectively. Particle Swarm Optimization (PSO) is one of the popular swarm based methods employed for optimizing digital circuits (Paul et al., 2015; Sasikumar and Muthaiah, 2017), and hence the same is studied to benchmark other optimization schemes. The PSO method employs swarm particles moving with a velocity while influenced by local and global optimal solutions. Swarm particles aid in moving to a globally best solution, whereas its cognitive behaviour helps in determining the particle’s (local) best solution. The social and cognitive components are constantly evolved to reach to a balanced positions, and thereby attain the optimized solution. Differential Evolution (DE) is known for achieving reliable results for real-world, non-linear optimization problems considering its competitive search accuracy, search robustness, and convergence speed. DE is not biologically inspired like other evolutionary approaches. However, like other evolutionary methods, the best candidate solutions are searched by applying mutation, crossover, and selection operators in the defined space. Multi-objective optimizer drives the circuit design by considering more than one hardware objective parameter at once instead of a single parameter or product of multiple parameters. A set of pareto-front solutions are deduced instead of a single solution when conflicting objective parameters are involved. In this work, five pairs of parameters derived from 8 single objective parameters were applied to evaluate dual objective algorithm. Similarly, four sets of three parameters derived from the original 8 parameters were applied for three objective algorithm. Non-dominated sorting genetic algorithm (NSGA) adopts the outline of genetic algorithm and applies modified selection of mating and survival (Srinivas and Deb, 1994). The best solutions are selected by combining parent and offspring populations with respect to fitness and spread. In NSGA-II, individual candidates are selected front-wise and continuously split based on crowding distance.

All six algorithms are distinctly different, with various control parameters to establish different dimensions within the large design space. Besides, the search vector progression is also not the same, considering different search strategies adopted by the algorithms and their effectiveness for different parameters. Hence it becomes challenging to develop intuition on algorithms adopted toward circuit optimization without putting the same into practice. This paper investigates various degrees of objectives along the selected parameters for optimizing the circuit of interest by employing six distinct optimization algorithms. All the techniques are inherently different and are likely to demand different computational requirements to reach similar solutions. Therefore, a finite number of 1000 evaluations composed of 100 generations/iterations and 10 candidate solutions at each generation/iteration with the same amount of transistor size change between consecutive generations is used to evaluate all the algorithm runs. The list of metrics adopted to evaluate different optimization techniques for custom circuit design across all the objectives are i) the best solution obtained, ii) monotonicity of fitness convergence along the generations, and iii) sparsity of candidate solutions.

3 OPTIMIZATION OBJECTIVES

Choosing the objective function to be minimised during optimization is important as it decides the circuit parameters trade-off. Hardware metrics in the form of delay, and different components of power, both individually and product of the two were considered for single-objective algorithm runs to optimize the design. Considering a positive correlation between area and power, only power is used in the objective function of single-objective algorithms. Following is the list of the hardware parameters which are configured as objectives to deduce the solutions from the algorithm runs: Delay (De), Leakage Power (LP), Switching-Power (SP), Sum of Switching-Power (SSP), Total Power (TP), Product of Delay and Switching Power (PDSP), Product of Delay and Sum of Switching Power (PDSSP), and Product of Delay and Sum of Total Power (PDSTP). Area as an ob-

![Figure 3: Schematic of a 28 Transistor Full Adder design.](image-url)
jective is employed for only multi-objective functions. The single objective and multi-objective derived design solutions are distinctly different and yet uniquely satisfy different design requirements. Since the set of algorithms are expected to work in large design space, a comprehensive evaluation is possible only for design with highest number of variables, transistor widths in this case. Hence, a 1-bit FA of 28 transistor design as shown in Figure 3 is considered suitable enough to check the efficiency of these algorithms. 28 variables also provide a large design space in the continuous domain, for which these metaheuristic optimization algorithms are expected to find solutions.

4 CIRCUIT MEASUREMENTS

The hardware parameters of the circuits were measured as depicted in Figure 4 (a). Timing, power, and area characterization was performed for the spice netlist given the PDK models, design rules, and PVT specifications. This work evaluated optimization methods by adopting a fast-fast process corner model from the Cadence gpdk 45 nm PDK, with $V_{DD}$ of 1 V, at a temperature of 70 °C. Other PVT conditions are expected to show a similar trend of optimization results. Circuit measurement of each candidate solution is performed in two stages: characterization and measurement. Characterization involves spice simulation of all possible timing arcs and covering all input vector combinations.

![Figure 4: Schematic representing (a) Circuit Measurement flow, and (b) generated Full Adder layout for circuit measurement.](image)

**Area**: The layout for candidate solutions is created using an experimental open-source tool Libre-cell (lib, 2022). The tool optimizes not only the cell width but also minimizes the wiring length by rendering optimal placement via the Eulerian path approach. The cell layout is limited to a fixed height between $V_{dd}$ rail at the top and Ground rail at the bottom. The MOS transistors are placed within this grid space by aligning the transistor’s width and length along the cell height and width dimensions, respectively. Higher drive strength cells are fit to the specified cell height by adopting folded MOS transistors. Additionally, shared diffusion to reduce parasitic capacitance is used along the Euler path. Post placement, the first two metal layers are used for internal cell routing. DRC, LVS rules of gpdk 45 nm files were satisfied by conservatively placing different entities such as diffusion, poly-gates, and metal contacts. The layout of FA design as generated from Libre-cell tool is shown in figure 4 (b).

**Delay**: The Non-linear delay model (NLDM) is obtained by exhaustively simulating all input-output timing arcs for all possible input vector combinations. Input transition from 20% to 80% and vice-versa in 6 ps to 300 ps is used to evaluate for two extreme input slew rates. Load capacitances of 0.1 fF and 70 fF are used to evaluate two extreme cases. For a different slew rate and load capacitance combination, measurement is then performed by using the NLDM model. Delay is measured as the time span between the input crossing 50% of the default supply value to the output crossing 50% level. The characterized delays for all possible input vector combinations are tabulated, and the worst case among all the characterized conditions is taken as the critical path delay.

**Power**: Different power terms are applied as objective goals, including switching-power, leakage-power, total-power, sum-of-switching-power, and sum-of-total-power. Power is also evaluated in two stages, like delay for extreme cases of input slew rate and output load. Power is measured by monitoring the current from $V_{dd}$ rail from the beginning of the input transition to the end of all the internal and output pin transitions reaching a level of 0.5% of the supply for falling transition, and 99.5% of the supply for the rising transition.

Switching power measures the rate of the energy dissipated by design under consideration when switching one or more inputs leading to the change in the output state while also accounting for the non-switching input and output pins, as they contribute to the total dissipated power for the state transition. The power is recorded for all the input vector combinations, covering all input-output arcs. For characterizing a custom circuit, the power dissipated during the charging-discharging cycle from the externally loaded
capacitor is deducted from the total power estimated from all the arcs. All possible input vector combinations are evaluated for state-dependent leakage power, and the average input vector combinations is treated as cell leakage power. The worst-case switching power among all the input-output arcs and input vector combinations is considered as the measured **switching power** (SP). This parameter, as an objective aims to establish reduced worst-case power when the circuit is employed for a system design. The **sum of switching power** (SSP) includes sum of switching power for all input-output arcs and input vector combinations. SSP parameter aids in establishing a circuit dissipating low average power, especially for a circuit where the occurrence of all input vector combinations have equal probability. **Total-power** (TP) refers to the cumulative sum of cell leakage power and the worst-case switching power. The **Sum-of-Total-power** (STP) represents the cumulative sum of SSP and LP.

5 RESULTS

The paper focuses on the computation and simulation effort to obtain quality circuits. All methods were evaluated for 100 generations and 10 candidates per generation. The setup was configured to run 8 targeted hardware parameters for single-objective algorithms, 5 pairs of hardware metrics for NSGA-II algorithm, and 4 sets of three hardware metrics for NSGA-III algorithm separately. Each algorithm was fed with the same 28T circuit topology and randomly generated transistor widths, while recording the candidate design solutions at every generation for further analysis. Individual algorithm control parameters were configured such that the amount of change in transistor size from one generation to the successive one remains the same.

5.1 Single Objective: GA, ES, PSO and DE

Figure 5 shows the best solutions from each generation, separately for eight single objective functions used. The four optimization algorithm aids in minimizing the selected parameters by varying the widths of the transistors at each generation. **GA** produces the best solution among the four algorithms. **DE** converges quickly to its best solution within 10 generations but does not offer the best hardware characteristics. All other algorithms converge to candi-
date solutions lower than the solutions generated by \textit{DE}, showcasing much fitter candidate solutions for a given number of evaluations. The candidate solutions reached by \textit{PSO} and \textit{ES} are comparable and consistently lie between the solutions generated from the other two algorithms. \textit{GA} and \textit{PSO} has a steady convergence rate. A non-zero falling convergence rate exists for \textit{GA}, \textit{ES} and \textit{PSO} at the end of the 100\textsuperscript{th} generation indicates room to yield fitter circuit designs. \textit{ES} shows monotonic behaviour for short durations before finding an improvement in fitness. It is to be noted that while \textit{GA} has higher convergence in simple single-objective functions, \textit{PSO} converges faster for product single-objective functions.

In terms of variation between candidate solutions in a generation, \textit{PSO} and \textit{DE} are observed to lose the diversity in solutions in the initial generations, as shown in Figure 6. \textit{PSO} is influenced by local optima of the best particle, and all particles in \textit{PSO} tend to move towards it. Hence \textit{PSO} hinders movement of other particles towards unexplored design space. \textit{ES} and \textit{GA} continue to carry good variation among the candidates to later generations, allowing them to explore a wider search space. \textit{DE} rapidly converges to a solution but fails to compete with solutions of the other three algorithms. Among all the objectives, delay and its product objectives, including \textit{PDSP}, \textit{PDSSP}, and \textit{PDSTP} showcase good variation among candidates of a generation, especially for \textit{GA} and \textit{ES} algorithms till the set 100 generations. A diverse set of solutions continue to be extracted from these two algorithms for delay and its product objectives. The other four objectives, which are related to power terms, demonstrate low diversity of candidates post the initial generations, which is expected since the power terms are directly related to the reduction in widths along the generation. Hence, linear progress on decreasing widths and lowering power is followed.

5.2 Two Objectives: NSGA-II

The \textit{NSGA-II} algorithm, a two-objective optimization method, was configured with following pair of objectives (in independent runs): i) Area and Delay, ii) Delay and Leakage Power, iii) Delay and Switching Power, iv) Delay and Total Power, and v) Delay and Sum of Total Power. The best candidate solution at each generation is presented in the top series of the plots in Figure 7, and all solutions are presented in the bottom series plot in the same figure. Within the 1000 evaluations composed of 100 generations and 10 candidates in each generation, it is difficult to find the optimal point as compared to the single objective functional run since the candidate solutions exhibit short-term fluctuations for the characterized delay. It is also clearly evident that apart from delay, all other parameters, including leakage power, switching power, total power, sum of total power, and area continue to showcase minuscule fluctuations along the generations. When the characterized delay shoots up,
the other objective parameters tends to dip and vice-versa. However, the algorithm tries to minimize delay in the further generations, along with the other objectives, either power or area. All candidate solutions in the plot show coherency for the power and area metrics but exhibits a large spread in the delay parameter. This is in accordance with the single objective study, where the delay and its product objectives continue to offer diverse solutions till the end. The best candidate solutions continue to present short-term fluctuations for the evaluated delay metric, so picking the right solution for a fixed generation/iterations run is challenging. One can pick a candidate design solution by adopting a moving average window along the generations for the delay. If the current evolved candidate circuit delay is lower than the moving average, then the evolution is terminated, and the current candidate solution is selected in the given evaluation frame. If otherwise, one can continue to search for an optimized solution. The moving average for a window length between 5 to 10 appears to be a good choice, considering runtime and quality of result.

5.3 Three Objectives: NSGA-III

Figure 8 shows the triple objective algorithm runs with the best solution in top row of plots, and all candidate solutions in the bottom row. NSGA-III is configured to optimize the following set of hardware metrics: i) Area, Delay and Switching power, ii) Area, Delay and sum of switching power, iii) Area, Delay and total power, and iv) Area, Delay and Sum of Total power. As expected, power and area follow a similar trend, but delay exhibits short-term fluctuations, as seen earlier. Hence, a moving average window aided termination for delay will help in selecting the best candidate circuit design after a threshold number of generations, as discussed previously. The three-objective runs reiterates the coherency of area and power parameters and the large spread in delay along the generations.

6 DISCUSSION

Figure 9 consolidates all the design solutions extracted from 6 different algorithms across all 100 generations, along with the pareto-front line. The candidate solutions that are among the pareto-front for the evaluated parameters are coloured in red in the plots of Figure 9. All 1000 solutions from 8 individual objective parameters based single-objective algorithms, 5 pairs of two objective parameters based NSGA-II algorithm, and 4 pairs of three objective parameters based NSGA-III are captured in the Figure. All the solutions are evaluated for area, total power, and delay. Figure 9 (a) cluster resembles almost a linear profile between power and area characteristics, as expected. Hence the optimum solution is the closest point to the origin for minimizing area and power, also marked in red colour. Figure 9 (b), and 9 (c) shows the cluster composed of all candidate solutions evaluated for delay versus area, and total-power versus delay, respectively. Single-objective algorithms offer the best solutions for a given objective, ignoring other hardware parameters. Nevertheless, NSGA-II and NSGA-III offers more candidate solutions that provide equal weightage to multiple hardware parameters. The multi-objective NSGA-II, and NSGA-III algorithm runs targeted toward multiple parameters are preferred over single-objective algorithm runs that are targeted towards the product of the same parameters. This is attributed to the domination of one objective over another in the single objective algorithm runs, which is not intended for optimizing towards two objective parameters. The red-coloured pareto-front line showcases the best solutions among the cluster of candidate solutions. One can select the best solution along the pareto-front line for the specified margin of the hardware parameters.
Figure 10: Extracted candidate solutions from all 100 generations and evaluated for total power versus delay with respect to (a) four single-objective algorithms covering all 8 objectives, (b) 8 different objectives configured on 4 different single-objective algorithms, (c) five pair of objective runs on NSGA-II algorithm, and (d) four set of objective runs on NSGA-III algorithm, (e) Single objective GA run for 8 objectives, (f) Single objective ES run for 8 objectives, (g) Single objective PSO run for 8 objectives, and (h) Single objective DE run for 8 objectives. Best viewed in enlarged size and color.

Figure 10 covers all solutions from different algorithms while segregating with respect to the objective function. Figure 9 (a) clearly shows that GA derived candidate solutions remain closest to the pareto-front line among all other single-objective algorithms when evaluated for characterized total power and delay. Figure 10 (b) showcases the impact of the choice of the single-objective function evaluated against Total-Power and Delay. Delay (De) and Product of Delay-and-Sum-of-Switching-Power (PDSSP) objectives resulted in candidate solutions with minimum delay. Switching-Power (SP), Sum-of-Switching-Power (SSP), and Product-of-delay-and-Switching-Power (PDSP) as objective parameters resulted in candidate solutions with low power. PDSP-based candidate solutions lie close to the pareto-front line, especially when both delay and total power are given equal weightage, as shown in Figure 10 (b). Single objective-based runs result in candidate design solutions highly optimized for either delay or power. Figure 10 (c) covers the candidate solutions generated from NSGA-II algorithm runs. As expected, the candidate solutions generated through the objective pairs consisting of delay and total power lie on the pareto-front points marked in red. Figure 10 (d) covers the candidate solutions generated from NSGA-III algorithm runs. The best candidate solutions generated through the objective pairs of Area, Delay, and Total-Power, lie on the pareto-front points. Additionally, Area, Delay, and Sum of Total-Power targeted solutions also lie on the pareto-front line, which is expected considering Sum of Total-Power involves worst-case Total-Power measurements.

Interestingly, the shape of the cluster consisting of candidate solutions generated from single-objective functions differs from the multi-objective derived ones when evaluated for the same hardware metrics. NSGA-II, and NSGA-III solutions move towards the origin with generations from the randomly seeded initial circuit. In contrast, the single-objective runs showed more significant movement towards the bottom and left, along the objective axis. Among the four single-objective functions, the derived pareto-front line composed of the best candidate solutions, from GA showcases the best hardware metrics, as depicted in the Figure 10 (e, f, g and h). It is also evident that delay and total-power targeted objective run showcases the best solutions when evaluated for the same hardware parameters for all four single-objective algorithms.

7 CONCLUSIONS

Circuit optimization is one of the common challenges encountered by designers. An ad-hoc usage of optimization algorithms has been employed without comprehensive evaluation. The proposed work evaluates...
six popular algorithms (GA, ES, DE, PSO, NSGA-II, NSGA-III) based on the impact from objective of optimization, sparsity presented by the evolved candidate solutions along the generations, diversity in hardware metrics of the candidate solutions, and monotonicity in convergence. All the algorithms were evaluated using FA design consisting of 28 transistors. Among the single-objective runs for the same number of evaluations, GA generates the most optimized circuit design solution and continues offering diverse solutions along the runs. ES comes close to GA and has a larger search space. PSO lacks diversity in solutions and is influenced by the best candidate in every generation. DE exhibits rapid initial convergence with poor hardware metrics. One can adopt GA algorithm over NSGA-II, and NSGA-III if the objectives and their importance are defined for the circuit under design. However, if the targeted parameter weights are not known or equal importance is recommended, then NSGA-II or NSGA-III are preferred over single-objective functions. This work aids in optimizing custom circuits and higher-order custom standard cells by adopting the most effective algorithm given the specified metrics range and evolution runs. The thorough investigation shows that a robust optimization of CMOS-based custom digital circuits is possible with a thorough characterization, irrespective of the technology progression, provided PDGs and interconnect models are made available.

REFERENCES


