Black-Box Optimization of Superconducting Circuits using Reduced-Complexity Neural Networks

Shrestha Bansal, *Student Member, IEEE,* Benjamin Chonigman, *Member, IEEE,* Chase Puglisi, *Student Member, IEEE,* Amol Inamdar, *Senior Member, IEEE,* Freddy Pena, *Member, IEEE,* Erik Lehmann, *Member, IEEE,* Subhanshu Gupta, *Senior Member, IEEE,* and Deepnarayan Gupta, *Fellow, IEEE*

Abstract—Single-flux quantum (SFQ) logic based high-speed periodic-threshold flash converter circuits require multiple non-linear and correlated parameters tuned precisely to function optimally. These parameters cannot be pre-determined from simulations and change with the clock frequency, thus requiring manual optimization every time the clock frequency is changed. In this work, we demonstrate automated bias optimization of an 8-bit superconducting SFQ flash ADC for the first time. A closed-loop 4X lower computational complexity hybrid particle-swarm-gradient-descent-optimization is demonstrated that first uses particle swarm optimization (PSO) to coarse tune the ADC biases and then applies gradient descent (GD) optimization for fine tuning. This results in a 12X reduction in the blind calibration time from several days manually, to a few hours using the proposed optimization scheme, with a resultant performance 2dB better than an optimization done by a highly skilled human.

Index Terms— Optimization, particle swarm optimization (PSO), gradient descent (GD), single flux quantum (SFQ), superconducting electronics, analog-to-digital converters (ADC).

I. INTRODUCTION

SUPERCONDUCTING electronic circuits deliver high switching speeds, high accuracy, low power, high sensitivity, and low noise. This makes these circuits well-positioned as enablers of future wireless technologies utilizing higher frequencies. Direct digitization of RF signals without the need to down convert them to a lower IF or baseband frequency domain is desirable to retain signal power and fidelity, which is beyond the scope of conventional silicon based data converters [1][2].

Operation of these circuits at cryogenic temperatures severely restricts them from achieving the high-level of integration that the conventional silicon-based circuits and systems can achieve, thus, restricting them to a select group of people having access to high-end laboratories and equipment. Silicon-based circuits are much more developed as compared to superconducting circuits and deliver remarkable computing performance required for backend digital tasks including calibration. Therefore, a hybrid system partly operating at room temperature and partly at cryogenic temperature is highly desirable. While extensive research is going on in the domain of superconducting circuits, there is still a need to perform system level enhancements to boost their performance and achieve their true potential.

Josephson Junctions (JJs) being two terminal devices require external control for optimal current biasing typically done in groups for the number of non-independent biases to be not too numerous. The absence of inherent gain in the JJ device and the highly correlated nature of bias currents in the existing designs thus makes it extremely time consuming to develop model-based bias optimization methodology. Bias optimization of mixed-signal superconducting circuits such as an analog-to-digital converter (ADC) is highly desirable because unlike the digital circuits, extracting the last few dBs of performance depends on the precise adjustment of a combination of multiple biases, making this problem a multi-parameter optimization problem. In state-of-the-art SFQ logic data converter circuits, this optimization is done manually which can take several days by an expert human to optimize the biases for a substantially new design. This calls for the need of an automated bias current optimization approach which can reduce the optimization time many folds, while improving the overall system performance.

This work will demonstrate a real-time closed-loop automated bias optimization of high-speed mixed-signal/digital JJ-based integrated circuits for the first time. The optimization methodology presented in this work treats the ADC as a black-box and uses the input-output characteristics to perform the system level optimization. To overcome fundamental limitations due to highly correlated bias currents in SFQ circuits, a multi-parameter optimization strategy is required that extends beyond the existing methods in prior art [3]-[5].

Neural network (NN) based optimization has been proven to optimize multiple parameters simultaneously at a fast rate [6]. While gradient descent (GD) optimization has been widely used to train NNs, it has a fundamental limitation of getting stuck in the local minima in the absence of a perfectly convex/concave transfer function [7]. Another stochastic optimization algorithm, particle swarm optimization (PSO), proposed by

Fig. 1. Proposed closed-loop optimization system architecture.

Fig. 2. Measured ADC spectrum with a ±10% offset in biases.

Eberhart [8] does not have this limitation of getting stuck in local minima as it works by spreading multiple particles in the search space and is thus a perfect choice for training the NN to optimize the SFQ logic based circuits [9]. PSO, however has a high computational hardware requirement [10].

Leveraging our recent modeling work on multi-parameter optimization techniques using artificial neural networks in [10], this work demonstrates a room temperature based closed-loop bias optimization methodology using neural networks for SFQ logic-based gigahertz data converters operating at 4K. In this work, a hybrid PSO-GD optimization algorithm first proposed in [10] by the authors is modified and experimentally validated. First the PSO coarse tunes the bias current values and locates the region of global minima, and then the batch GD algorithm is invoked to fine tune the bias currents and locate the region of global minima. The batch GD replaces the stochastic GD used in [10] for better convergence.

The neural-network based optimization system architecture is shown in Fig. 1. It works in a closed-loop with the 8-bit SFQ logic based flash ADC. Inputs to the neural network are the ADC biases which are internally fed back and the difference between the expected ADC output ($y_i[k]$) and the practical ADC output ($y[k]$) as the error function (e_i).

The following key additions and modifications are presented in this work over the initial work on algorithm modeling and development in [10]:

1) Review of the 8-bit SFQ ADC and recent state-of-the-art optimization techniques for SFQ circuits (Section II)

2) Analysis of the proposed hybrid PSO-GD optimization algorithm highlighting key differences from [10] (Section III)

3) Detailed description of the closed-loop ADC system including design, optimization, and measurement setup with the ICE-T and the external NN controlled bias optimizer at room temperature (Section IV), and

4) Measurement results with detailed discussions on the improvements achieved using the proposed optimization methodology. (Section V)

We conclude this work in section VI including scope of possible future research.

II. SFQ ADC AND STATE-OF-THE-ART OPTIMIZATION TECHNIQUES

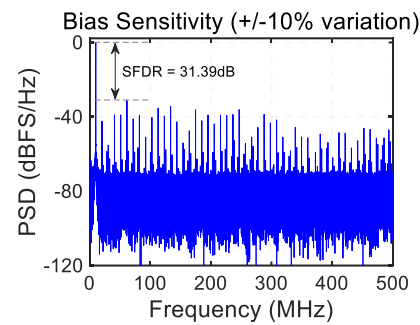Superconducting flash ADC was chosen as an exemplar superconducting mixed-signal circuit due to the complexity and difficulty of its optimization. It uses comparators that exploit the periodic nature of the superconducting quantum interference devices (SQUIDS), and therefore exhibit multiple comparison thresholds leading to a significant reduction in hardware per ADC as compared to their silicon counterparts [11]. These comparators digitize the periodic current flowing through the quasi one-junction SQUID (QOS) junction. All such comparators exhibit dynamic distortions, which are manifested as a non-sinusoidal asymmetric transfer function. These dynamic distortions are a result of the phase-dependent QOS junction inductance.

The flash ADC used as a testbed for black-box optimization in this work is derived from [2] and uses a DQOS logic-based comparator to significantly reduce the non-linearities and has Gray coded output bits. It is designed such that each higher significant bit is precisely positioned at the center of its preceding bit. As observed in our experiments, these bits do not inherently switch at the exact center point of their preceding bit, thus leading to errors in the reconstructed signal waveform. This behavior of the bits is controlled by three bias controls for each comparator, making a total of twenty-four biases for the 8-bit ADC. These biases are: the differential phase bias, the duty cycle bias, and the threshold phase bias. Differential phase biases and duty cycle biases control the respective differential phases and duty cycles of the gray coded output and do not significantly contribute to the alignment of the output bits. Threshold phase bias on the other hand controls the trip point of each bit and hence directly contributes to the alignment of these output bits, hence making it most crucial to obtain the accurate reconstruction of the input signal. As shown in Fig. 2, due to this high sensitivity of threshold phase biases, even a +/-10% variation in these biases from the manually optimized point reduces the SFDR performance by more than 20dB. These biases are dependent on the biases of the previous bit, with the MSB bit bias having the maximum effect and LSB bit bias the least effect. This interdependence of the biases makes the optimization by manual tuning a difficult task. For the scope of this work, we manually optimize the two lesser sensitive biases which have minimal effect on the output waveform and use the devised algorithm to optimize the most sensitive threshold phase biases.

Several digital calibration techniques exist for optimization of silicon-based circuits and systems, which enable a hardware-software co-design to improve their performance. Prior works on the optimization of SFQ circuits have been limited to the optimization of circuit parameters during the design phase of
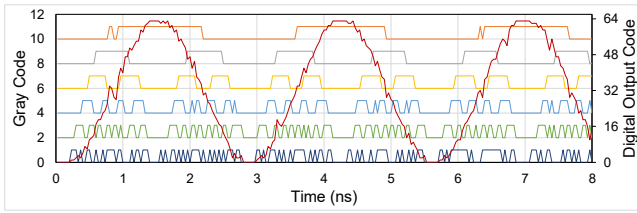
Fig. 3. Phase mismatch results in misaligned gray codes causing multiple glitches during signal reconstruction as shown for the 6 MSB output bits.

the circuits. For improving the yield, the method of inscribed hyperspheres [12] and the center of gravity method [13] have been widely used for optimization. These methods are however computationally expensive. The critical margin method [14] is widely used for improving the current margin and is computationally less expensive. However, like GD, the critical margin method also gets trapped in local minima easily. Monte-Carlo based optimization of circuit parameters that analyze the effect of process variation has been extensively used by circuit designers across the silicon and superconducting domains [15].

However, to the best of our knowledge, prior art does not address system level optimization of SFQ logic based circuits such as data converters (ADCs). These ADCs are currently manually optimized by a skilled professional which is not only time exhaustive but also inefficient and costly. It is thus important to investigate optimization methodologies working at room temperature (for ease of control) that can efficiently optimize the bias currents feeding the superconducting circuits at cryogenic temperatures with a lower closed-loop computational and integration complexity.

III. HYBRID PSO-GD OPTIMIZATION ALGORITHM

The interdependence of the bias values of the SFQ ADC that this work utilizes significantly complicates the optimization task. Even though the broad range of these bias values is available from the simulations, their exact values are not. In the absence of these exact bias values, ADC output is completely distorted, and thus requires a) start-up of the ADC, and b) calibration of the biases to get the optimum performance.

The hybrid PSO-GD algorithm first presented in [10] is a computationally efficient and low latency algorithm which has shown merit in performing multi-parameter optimization of silicon-based ADCs. The PSO-GD algorithm in [10] utilizes PSO to coarse tune the ADC parameters and then uses stochastic gradient descent (SGD) to fine tune the ADC parameters, achieving global minima with lesser number of computations. In this work, we modify the originally presented PSO-GD algorithm by replacing SGD with batch gradient descent (BGD). For optimizing k parameters, PSO works by spreading n-particles in a k-dimensional search space, with each particle working in collaboration and competition with the other particles to achieve the global optimum solution. The update equations for PSO are shown in (1) and (2), where (1) defines the particle velocity and (2) defines the particle position.

$$v_i^{t+1} = w \cdot v^t + c_1 \cdot \text{rand}(0,1) \cdot \left\{ x^t_{p_{best}} - x^t \right\}$$
$$+ c_2 \cdot \text{rand}(0,1) \cdot \left\{ x^t_{g_{best}} - x^t \right\} \quad (1)$$
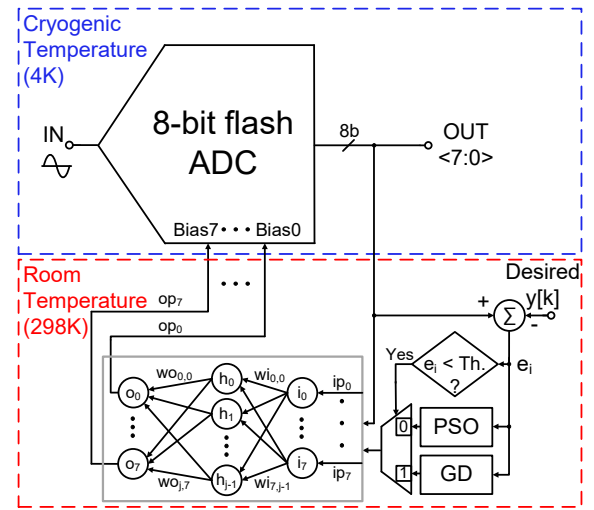

Fig. 4. Closed-loop optimization system architecture for heterogeneous temperature optimization.

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (2)$$

The first term in (1) represents the inertia factor which directs the particle to move in the direction of inertia from the previous iterations. The second term in (1) represents the competition factor, which directs the particle to compete with its own personal best. The third term represents the cooperation factor, which dictates the particles in the swarm to assist each other to locate the global best position corresponding to the global minima. Vector addition of the updated velocity v_i^{t+1} from (1) is then performed with the particle's previous position x_i^t to arrive at the next location x_i^{t+1}. These n-particles are randomly initialized using a random number generator in the expected optimization space provided by the circuit designer from the simulations. This optimization space is the range in which a particular parameter can be optimally located and is shared by one dimension (corresponding to that parameter) of each particle. Thus, for a n-particle k-dimensional system, we have k individual optimization spaces in which the global optimum solution can be found, making it a k-dimensional optimization problem. After a series of iterations, when the error function is reduced to 5% of the desired value, the best position from the set of n-particles is used to as the starting position of GD. This 5% handoff criterion strikes a balance between the optimization time and computational complexity [10]. Further, SGD updates the positions of these k-parameters to arrive at the global optimum solution using the update rule in (3) as follows:

$$x^{t+1} = x^t - 2 \cdot \mu \cdot \frac{1}{m} \sum_{i=1}^{m} (e^m) \quad (3)$$

where x^{t+1} represents the updated position, x^t represents the previous position, μ represents the weight factor, and the error function e is summed for all the m samples in the dataset.

SGD uses a single training data sample at a time to perform optimization while traversing through the entire dataset, making it computationally fast. However, the frequent weight updates happening at each subsequent data sample, cause the update function to be noisy and may cause oscillations in the error function as a result. It also consumes more power to perform the optimization since the entire computing hardware
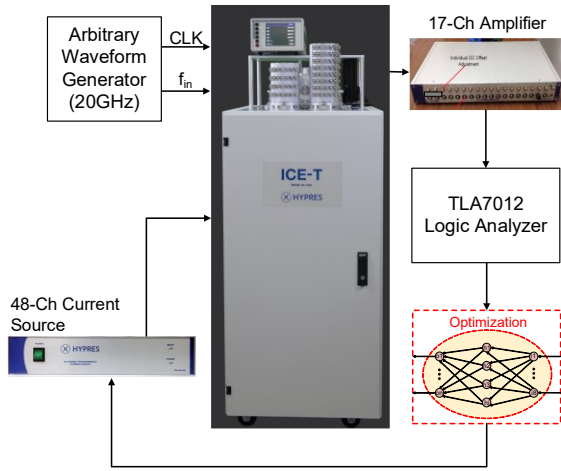
Fig. 5. Integrated cryogenic electronics testbed with room-temperature based optimization.

IV. SYSTEM DESIGN AND MEASUREMENT SETUP

Fig. 4 shows the optimization system architecture with a 10.3MHz sinusoidal input, IN, which is fed to the ADC to perform the conversion. The digital output, OUT, is first synchronized with the digitally sampled input signal, y[k], and then compared with this digital output to obtain an error signal, e_i. Initially, the NN with the input layer weights w_i and output layer weights w_o, is trained using PSO algorithm and when e_i reaches below the specified 5% threshold (e_{ref}), BGD takes over the optimization task. In this work, since the eight threshold phase biases are optimized, each PSO particle produces a set of eight outputs ranging between 0-3mA that feed the eight threshold phase biases (Bias0 - Bias7). The global best particle position from PSO is then used by the BGD to further optimize the system.

The measurement setup for the optimization of the 8-bit flash ADC is shown in Fig. 5. The ADC is mounted on a universal 80-coax insert for superconductor chips and is housed in a Hypres Integrated cryogenic electronics test-bed (ICE-T) which precisely sets the ambient temperature of the chip at 4K. Connections to the ADC are made via these inserts having SMA connectors which directly connect with the peripheral devices at room temperature. Clock and signal inputs to the ADC are provided via a 20GHz arbitrary waveform generator. A 48-channel Hypres CS-48-100 precision current source feeds the eight sets of three biases (one set for each comparator) and another set of biases which control the GPIOs and the enable circuitry of the flash ADC chip. Output of the ADC is fed to a Hypres 17-channel amplifier which enables the reading of the output by the Tektronix TLA7012 logic analyzer by amplifying the output waveforms to a full-scale voltage level. The logic analyzer output is then processed in MATLAB and Python where the optimization algorithm analyzes this data and modifies the bias current values to perform the optimization. This in turn closes the feedback loop as in Fig. 5.

is used to perform updates for every sample. Another issue with SGD is that in case of a transfer function where one sample does not linearly depend on the previous sample, that is, when there are multiple random spikes in the dataset as in the case of SFQ ADCs (shown in Fig. 3), the updates for the samples with large deviation from the general trend may prevent the system from converging, which is highly undesirable [16]. A solution to this problem is using BGD, which although slower than SGD, takes the entire dataset into consideration while computing the error function. This results in a more stable gradient of convergence as compared to SGD. Moreover, it performs vector operations and can utilize parallel processing of the microprocessor, thus reducing the optimization time significantly. Thus, the bias currents are first optimized by using PSO to find the region in the vicinity of global minima, and BGD is subsequently used to fine tune the biases to converge to the global minima faster and with greater accuracy.
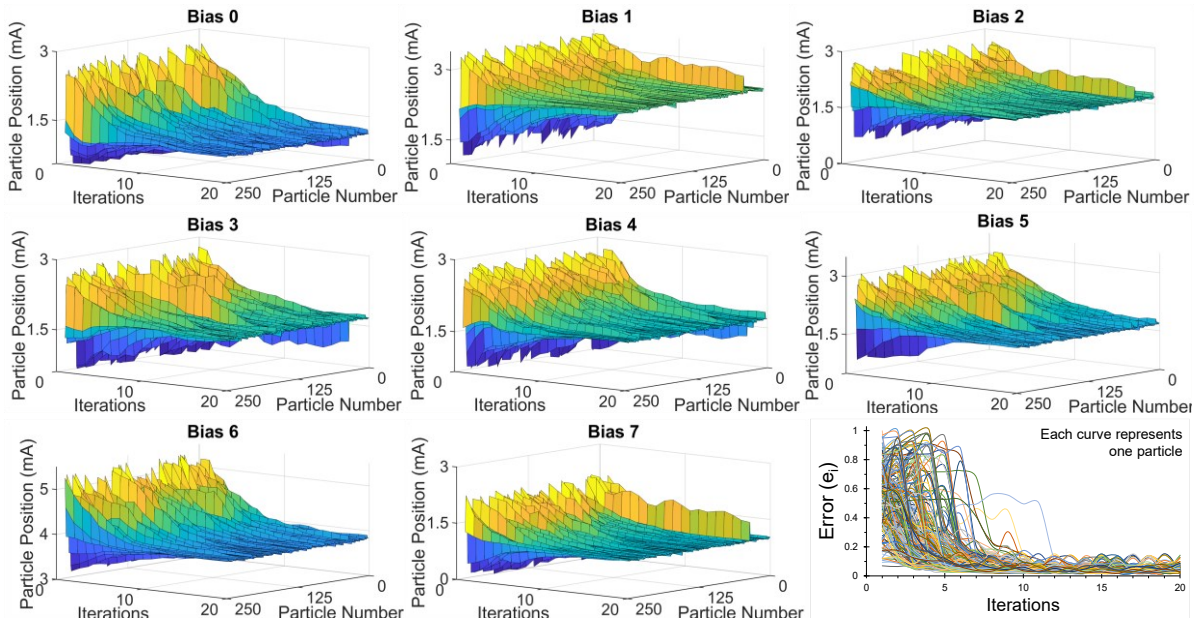

Fig. 6. Convergence of the 8 threshold bias values over 20 iterations and the error function against number of iterations for 250 particles (bottom).
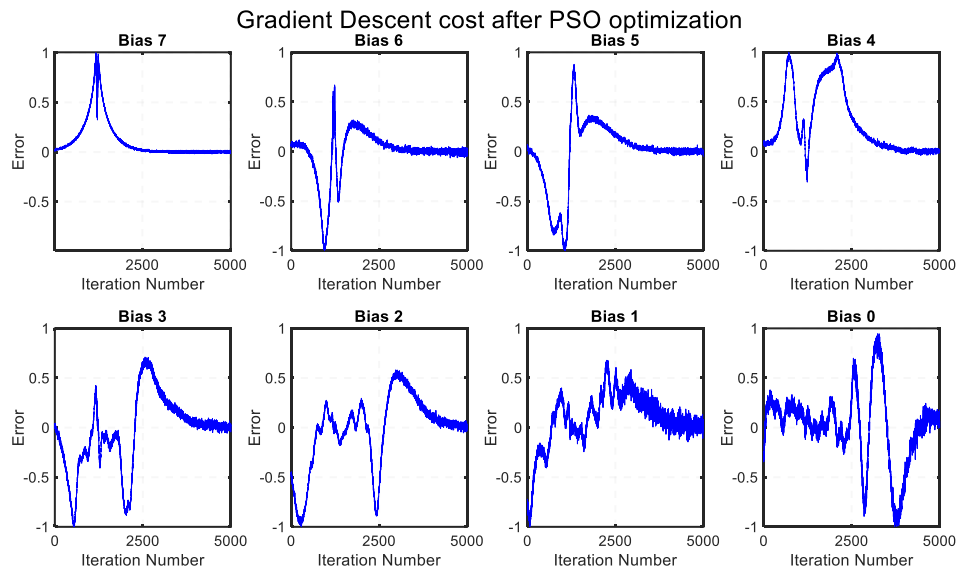
Fig. 7. Measured error functions (in mA) showing the cost function for the 8 ADC biases converge using BGD to fine-tune the ADC output.

The optimization algorithm requires precise alignment of the input signal with the reconstructed ADC data to perform optimization. However, due to the inherent delay of the various systems involved, it is impossible to get the input signals aligned with the reconstructed output data, rendering the optimization algorithm in need of a synchronizer. To implement this synchronizer module, the algorithm is modified in a way to detect the peaks of the output data and the input signal. A fixed number of data points are then captured from both the waveforms starting from these peaks. This solves the synchronization issue and enables the algorithm to optimize the biases in real-time by comparing each sample of the output data with the quantized input signal.

V. MEASUREMENT RESULTS

The flash ADC was designed and fabricated in the MIT-LL 10kA/cm² process. Due to the hardware constraints in the data acquisition, measurement results have only been verified at an input signal frequency (f_{in}) of 10.3MHz with the sampling clock frequency set at 1GHz. Using FPGA to acquire the ADC output is an alternate method that can support clock frequencies upwards of 10GHz, but for the scope of this work, it has been excluded.

After the eight differential phase biases and eight duty cycle biases are manually optimized, the eight threshold phase biases are randomly initialized for optimization using the proposed algorithm. Starting with these randomly initialized values, PSO works to locate the region of global minima. When the error between the quantized input and the ADC output data is reduced to 5%, the PSO hands over the control to the BGD for fine tuning of the biases. PSO utilizes a particle which is a 1x8 vector. Each dimension of this vector represents the value of one threshold phase bias between 0 to 3mA. Using a group of 250 such particles, PSO converges to the vicinity of the global minima in 20 iterations as shown in Fig. 6. Next, the particle position providing the best convergence amongst all the particles obtained after these 20 iterations of PSO is chosen as the starting point of the BGD which then takes over the optimization task to converge to the global minima with a greater accuracy. The convergence of the BGD is as shown in Fig. 7. Lower significant biases (Bias 0, Bias 1 and so on in Fig. 7) are observed to have a greater deviation from the mean position, which is because the effect of the biases on the overall error function decreases as the bit significance goes from high to low.

Using only the conventional PSO to optimize the ADC, a spurious free dynamic range (SFDR) value of 53.81dB is achieved in 20 iterations with 1000 particles. On the other hand, using the presented hybrid optimization algorithm, we reduce the number of particles required to perform coarse optimization by PSO to 250 particles, and it is observed that in 20 iterations
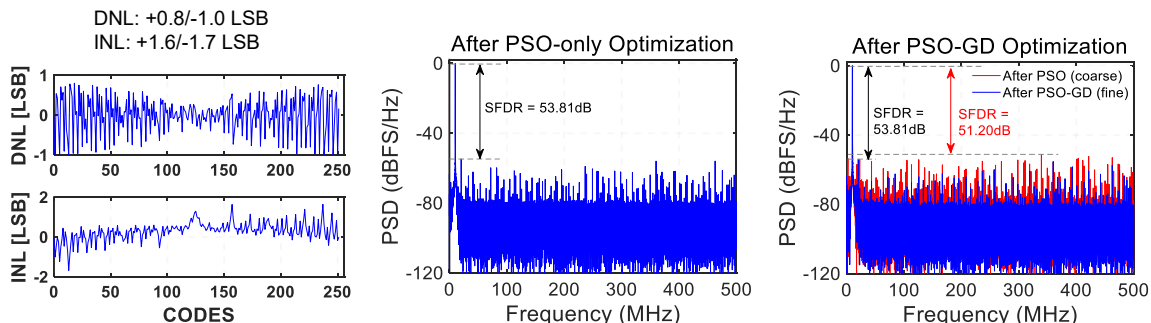

Fig. 8. Measured INL/DNL (left), Measured ADC spectrum after PSO-only (with 1000 particles) (middle) and after proposed PSO-BGD optimization (right).

TABLE I
COMPARISON OF HARDWARE COMPLEXITY AND PERFORMANCE OF THE DEVISED PSO-BGD ALGORITHM WITH PSO-ONLY

Conditions		SFDR improvement[†] (dB)	SNDR improvement[$][†] (dB)	Additions	Multiplications	ADC Calls
With a +/-10% bias variation from manually tuned values		-20.42	-12.21	-	-	-
Optimization using PSO only (n1 = 20, n2 = 1000)		2	1.80	200M	100K	20K
Proposed (PSO-GD) Optimization (n1 = 20, n2 = 250, lgd = 5000)	PSO (Coarse)	-0.61	-1.63	50.12M	40K	10K
	GD (Fine)	2	2.41			

n1: Number of PSO iterations; n2: Number of swarm particles; ls: Number of samples (5000); lgd: Number of iterations for GD using proposed algorithm, n: Number of ADC output bits (8)

For PSO only, #Additions: $n1 * n2 * 2l_s$, #Multiplications: $5 * n1 * n2$, #ADC Calls: $n1 * n2$
For the PSO-GD, #Additions: $n1 * n2 * 2l_s + 2n*l_s + 8*l_{gd}$, #Multiplications: $5 * n1 * n2 + 3 * l_{gd}$, #ADC Calls: $n1 * n2 + l_{gd}$

[$]Integrated over 500MHz, for f_{clk} = 1GHz and f_{sig} = 10.3MHz
[†]Reference is the best values obtained with manual calibration

a 51.20dB SFDR performance is achieved which after fine optimization by BGD further improves to 53.81dB (Fig. 8), making it identical to the performance achieved using only PSO based optimization, but with reduced computational complexity. This SFDR value is 2dB better than the best SFDR performance achieved by manual optimization of the biases. Similar improvements are also seen for SNDR. The relative SNDR performance improvement is observed to be 1.8dB better than the best achieved with manual optimization which further improves to 2.41dB using the hybrid algorithm. A 1.6LSB/0.8LSB INL/DNL performance is achieved post optimization using the hybrid algorithm.

Table I shows the performance comparison between the presented hybrid optimization algorithm and the PSO-only optimization method. The hybrid PSO-GD optimization algorithm using 250 PSO particles reduces the number of ADC calls required to optimize by 50%, multiplications by 60% and addition operations by 75% as compared to the PSO only optimization method that requires 1000 particles to achieve identical performance. This leads to an overall 12X reduction in optimization time, with a lower computational complexity over the conventional PSO optimization.

VI. CONCLUSION AND FUTURE WORK

In this work, a room temperature based bias optimization technique for superconducting circuits is demonstrated and validated on an 8-bit SFQ flash ADC. The hybrid PSO-BGD optimization algorithm reduces the optimization time by 12X while reducing the total number of computations by 4X. A 2dB relative SFDR improvement with a 2.41dB relative SNDR improvement versus the best human performance obtained by manual tuning is observed. To our knowledge, this is the first work that uses classical optimization techniques operating at room temperature to optimize superconducting circuits operating at cryogenic temperature. While the SFQ flash ADC was chosen as a testbed for demonstrating the effectiveness of the devised optimization methodology, it can be used to optimize any other superconducting system since it treats the system as a black-box. This work thus opens a new horizon for superconducting circuits by enabling the high level of integration seen in conventional silicon circuits while realizing their full potential, thus significantly contributing to improving the testing and characterization of superconducting circuits.

REFERENCES

[1] D. Gupta, et al, "Superconductor analog-to-digital converters and their applications," IEEE MTT-S Intl. Micro. Symp., Balt., MD, 2011, pp. 1-4.
[2] A. Inamdar, et al, "Design and evaluation of Flash ADC," in IEEE Tran. on Appl. Super., vol. 25, no. 3, pp. 1-5, Jun. 2015.
[3] Takashi Oshima, et al, "Fast nonlinear deterministic calibration of pipelined A/D converters," IEEE Mid. Symp. on Cir. and Syst, (MWSCAS), 2008, pp. 914–917.
[4] Un-Ku Moon, et al, "Background digital calibration techniques for pipelined ADCs," IEEE Trans. on Cir. and Syst. II: Analog and Digital Signal Processing, vol. 44, no. 2, pp. 102–109, Feb. 1997.
[5] S. Gupta, et al, "Multi-rate polyphase DSP and LMS calibration schemes for oversampled data conversion systems," IEEE Intl. Conf. on Acoustics, Speech and Signal Proc. (ICASSP), 2011, pp. 1585–1588.
[6] P. D. Reynolds, et al, "FPGA implementation of particle swarm optimization for inversion of large neural networks," Proc. of the IEEE Swarm Intell. Symp. (SIS'05), 2005, pp. 389–392.
[7] B. Widrow, et al, "A comparison of adaptive algorithms based on the methods of steepest descent and random search," IEEE Trans. on Ant. and Prop., vol. 24, no. 5, pp. 615–637, Sep. 1976.
[8] J. Kennedy and R. Eberhart, "Particle swarm optimization," IEEE Intl. Conf. on Neu. Netw., 1995, pp. 1942–1948.
[9] D. Parrott and Xiaodong Li, "Locating and tracking multiple dynamic optima by a particle swarm model using speciation," IEEE Tran. on Evol. Comp., vol. 10, no. 4, pp. 440-458, Aug. 2006.
[10] S. Bansal, et al, "Neural-Network Based Self-Initializing Algorithm for Multi-Parameter Optimization of High-Speed ADCs," IEEE Tran. on Cir. and Syst. II: Express Briefs, vol. 68, no. 1, pp. 106-110, Jan. 2021.
[11] D. Petersen, et al., "A high-speed analog-to-digital converter using Josephson self-gating-AND comparators," IEEE Tran. on Mag., vol. 21, no. 2, pp. 200-203, March 1985.
[12] Q. P. Herr and M. J. Feldman, "Multiparameter optimization of RSFQ circuits using the method of inscribed hyperspheres," IEEE Tran. on Appl. Super., vol. 5, no. 2, pp. 3337-3340, June 1995.
[13] T. Harnisch, et al, "Design centering methods for yield optimization of cryoelectronic circuits," IEEE Tran. on Appl. Super., vol. 7, no. 2, pp. 3434-3437, June 1997.
[14] C. A. Hamilton and K. C. Gilbert, "Margins and yield in single flux quantum logic," IEEE Tran. on Appl. Super., vol. 1, no. 4, pp. 157-163, Dec. 1991.
[15] M. Jeffery, et al, "Monte Carlo optimization of superconducting complementary output switching logic circuits," IEEE Tran. on Appl. Super., vol. 8, no. 3, pp. 104-119, Sept. 1998.
[16] L. Bottou, "Online learning and stochastic approximations," On-line learning in Neu. Net., Cambridge UP, 1998.