Heuristic Model Structure Optimization for Digital Predistortion

WEN QIAO¹, HONGMIN LI², CHENGYE JIANG³, GUICHEN YANG⁴, LEI SU⁵, AND FALIN LIU⁶

Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China
Key laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei 230027, China

Corresponding author: Falin Liu (liuf@ustc.edu.cn).

This work was supported by the National Natural Science Foundation of China under Grant 61471333

ABSTRACT Power amplifiers (PAs) are widely used in RF broadcasting applications. However, they exhibit nonlinear behavior and deteriorate the quality of transmitted signals. Digital predistortion (DPD) is developed to linearize the distortion generated by PAs. Due to the contradiction between modeling precision and complexity, to reduce the complexity and ensure comparable performance, it is necessary to determine an appropriate or optimized structure before applying DPD. Heuristic method is a good method to solve such multivariate problems with a considerable large search domain. In this paper, a novel approach to determine the structure of DPD based on heuristic method is proposed, including the enhanced hill-climbing (EHC) with stronger global search ability and enhanced genetic algorithm (EGA). Ridge regression is introduced to ensure the correctness of the search direction. Besides, orthogonal matching pursuit is used to further reduce the basis number while keeping good performance. The validation of the proposed techniques on a Doherty PA using generalized memory polynomial (GMP) model demonstrates the capability to efficiently find the dimensions of an appropriate GMP model. As shown by the convergence curve and the final model performance, the model searched by the proposed method has satisfactory results, and achieves a good balance between complexity and performance.

INDEX TERMS Digital predistortion, genetic algorithm, heuristic, hill-climbing, orthogonal matching pursuit, power amplifiers, ridge regression.

I. INTRODUCTION

Driven by the combination of market demands and technology advances, the next generation communication system, i.e., 5G, had become one of the research hotspots and been commercialized in 2020 [1]. The signal transmitted in the system is characterized by ultra-wideband and a high data transmission rate [2]-[4]. The power amplifier (PA) is one of the indispensable components in the system and it is also the major component causing nonlinearity, which deteriorates the quality of the transmitted signal both in time and frequency domain.

Therefore, the behavioral modeling and linearization of PAs have been widely studied. Many methods have been proposed to alleviate the distortion, such as power back off (BO), analog predistortion (APD), and digital predistortion (DPD) [5] [6]. DPD attracts wide attention among these methods, owing to its flexibility and high efficiency.

Volterra series [7] is typically adopted to characterize nonlinear systems, which has also been introduced into PA modeling. However, due to the high correlation of Volterra series, too many basis functions have to be included when characterize a PA with high nonlinearity. Therefore, many simplified versions have been proposed, such as memory polynomial (MP) model [8], generalized memory polynomial (GMP) model [9] and dynamic deviation reduction-based (DDR) model [10]. As the bandwidth of the transmitted signal becomes increasingly wider, MP model is not accurate enough to modify the PA. Thus, GMP is more attractive because of its capability to model a system with relatively strong nonlinearity. The structure of these models is determined by the nonlinear orders and the memory depths, representing the static and dynamic nonlinearity, respectively. Without loss of generality, GMP is the model used in this article, and the proposed search algorithm can also be applied to other models such as MP, DDR, etc.

Before applying DPD to linearize PAs, it is necessary to first determine the structure of the model which is suitable for corresponding scenarios.
Basically, the requirements for modeling performance and modeling complexity are contradictory. An appropriate model structure can make a good tradeoff between the computational complexity and the performance [11]. Classically, the model structure is set with the corresponding parameters chosen by experience, which makes it difficult to balance model performance and complexity. In other words, either an over-fitting or an under-fitting is likely to appear and deteriorate the model performance.

Therefore, it is necessary to solve this problem. Exhaustive search can naturally find the best structure of a model while it is too computation costly to implement in a real system. Fortunately, heuristic search guides the search by the heuristic information of the problem to reduce the complexity and the search domain corresponding to the problem, where hill-climbing (HC) method [12], genetic algorithm (GA) [13], simulated annealing (SA) [14] and particle swarm optimization (PSO) [15] are classical algorithms. Therefore, in this paper we propose a method based on heuristic method to determine the proper structure of GMP model.

However, even in the simplified models, such as MP, GMP, DDR, basic functions contribute differently to the modeling performance. In [16], a non-uniform memory polynomial behavioral model is proposed to further reduce the number of parameters in the MP model without affecting its performance. A method based on the attention mechanism is used to choose the memory terms effectively in [17]. Similarly, basis functions with different nonlinear orders also contribute differently to the modeling performance. In [18], by ignoring the nonlinear terms that contribute less to the modeling performance at the same memory depth, an MP model with non-uniform nonlinearity is proposed. The phenomena mentioned above are attributed to the correlation between basis functions of models derived from the Volterra series. Unfortunately, in [13]-[15], model structures obtained by search algorithm based on simplified Volterra-like model still have redundancy. If only the model structure is searched, basis functions redundancy cannot be avoided.

Therefore, it is necessary to select an appropriate subset of the searched models. At the same time, the selected subset can achieve a comparable modeling performance with the searched structure and reduce the modeling complexity further. Orthogonal matching pursuit (OMP) contributes to picking the primary proper subset in an efficient way [19].

In this paper, search methods based on heuristic are proposed to search DPD model structures, where the fitness value is utilized to generate the decision of solutions in the next step and converges quickly. The proposed methods also possess the following advantages: multiple initial points search in parallel, quick convergence, enhanced numerical stability in the search process to ensure a correct iterative search direction, and accepting a poor solution in some subdomains with a certain probability which may contribute to jump out of the local optimal solution and finally reach the global optimal solution.

The remainder of this paper is organized as follows: Section II introduces the DPD background, heuristic background, and the definitions of the heuristic algorithms. In section III, novel methods based on the heuristic algorithm and appropriate implementation structure are proposed to identify the appropriate DPD model structure. Experimental results are presented in section IV. In section V, a conclusion is drawn.

II. HEURISTIC STRUCTURE SETTING

A. DPD BACKGROUND

In terms of model identification, indirect learning (IDL) [20] and direct learning (DL) [21] are classic structures. In this paper, the indirect learning architecture (IDL) [20] is considered and shown in Fig. 1. One of the characteristics of IDL is the quick convergence capability. In order to effectively linearize the nonlinearity of PAs in broadband scenarios, the GMP model is adopted as an example, which is represented as:

$$
\begin{align*}
  y_{GMP}(n) &= \sum_{k=0}^{K_a-1} \sum_{l=0}^{L_a-1} a_{kl} x(n-l)|x(n-l)|^k \\
  &+ \sum_{k=1}^{K_b} \sum_{l=0}^{L_b-1} \sum_{m=0}^{M_b} b_{klm} x(n-l)|x(n-l-m)|^k \\
  &+ \sum_{k=1}^{K_c} \sum_{l=0}^{L_c-1} \sum_{m=1}^{M_c} c_{klm} x(n-l)|x(n-l+m)|^k
\end{align*}
$$

(1)

where $K_a$ and $L_a$ are nonlinear order and memory depth of the aligned signal; $K_b$, $L_b$ and $M_b$ are the memory depth, nonlinear orders, and cross terms length of the lagging term; $K_c$, $L_c$ and $M_c$ are the memory depth, nonlinear orders, and cross terms length of the leading term.

Represent (1) by a matrix as

$$
  y = X\alpha,
$$

(2)

where

$$
  \alpha = [a_{kl} \ b_{klm} \ c_{klm}]^T,
$$

(3)

$$
  X = [X^{(1)} \ X^{(2)} \ X^{(3)}],
$$

(4)

where $a_{kl} \in \mathbb{C}^{K_a \times 1}$, $b_{klm} \in \mathbb{C}^{K_b \times 1}$ and $c_{klm} \in \mathbb{C}^{K_c \times 1}$ stand for the coefficients and $X^{(1)} \in \mathbb{C}^{N \times K_a}$, $X^{(2)} \in \mathbb{C}^{N \times K_b}$, $X^{(3)} \in \mathbb{C}^{N \times K_c}$.
The basis functions corresponding to aligned signal envelop, lagging envelope and leading envelope, respectively. $K_1$, $K_2$, and $K_3$ are the number of parameters of aligned, leading, and lagging signals, $N$ is the sample length.

The cost function can be represented as

$$\alpha = (X^H X)^{-1} X y$$

where $\alpha \in \mathbb{C}^{K \times 1}$ represents the parameters of the post-inverse of the PA, $K = K_1 + K_2 + K_3$ is total number of parameters and $H$ represents conjugate transpose. If a system can be modeled by a pth order model, according to [22], the pth pre-inverse is equal to the pth post-inverse of the model. Correspondingly, by swapping the input $x$ and output $y$ in the above procedure, the DPD model is constructed. The DPD coefficients can be obtained and expressed as follows

$$\alpha_{DPD} = (Y^H Y)^{-1} Y x.$$  

\section*{B. BASIC FACTORS OF HEURISTIC ALGORITHMS}

The neighborhood is one of the main conceptions in optimal domain, which is defined in terms of Euclidean distance:

$$U^\delta(n): 0 < \sum_{i=1}^{N_p} (p_m(n) - p_m(n))^2 \leq \delta$$

where $U^\delta(n) \in \mathbb{R}^{N_p \times N_b}$ is the neighborhood of $p_m(n) = [p_1(n), ..., p_N(n)]$ in the nth iteration, $N_p$ is hyper parameters of DPD model. $U_\delta(n)$ includes $N_b$ neighbors satisfying (8). $\delta > 0$ stands for the Euclidean radius of the current point. The search point is defined as a coordinate with 8 dimensions in GMP model ($N_p = 8$), which specifically stand for the eight parameters of GMP model denoted as $p_i : (k_a, L_a, k_b, L_b, m^l_b, k^l_b, m^L_b)$, and $p_i \in \mathbb{R}^{8 \times 1}$.

The merit to measure model performance is the normalized mean square error (NMSE), which denotes as

$$NMSE = 10 \log_{10} \left( \frac{\sum_{n=1}^{N} |x(n) - y(n)|^2}{\sum_{n=1}^{N} |x(n)|^2} \right)$$

where $N$ stands for the length of the input signal, $x(n) \in \mathbb{C}^{N \times 1}$ is the input signal, $y(n) \in \mathbb{C}^{N \times 1}$ is the output of the PA after DPD.

Due to the complexity of coefficients extraction directly relates to parameter number, the complexity of the model can be measured by the number of parameters of the GMP which is marked as $N_{coeff}$. Considering both modeling performance and complexity, a cost function can be proposed as:

$$J_i = NMSE_i + \beta f(N_{coeff}(i))$$

where $J_i$ and $NMSE_i$ represent the $i$th fitness value and performance evaluated by NMSE. $f(N_{coeff}(i))$ is a function of $i$th parameter number $N_{coeff}(i)$, $\beta$ is the regularization factor. The trade-off between the performance and the complexity of the model is governed by $\beta$, which is an indicator of how much importance the complexity of the model is. So far, it can be concluded that what needs to be solved is a multi-objective optimization problem, specifically the optimization of both performance and complexity.

\section*{III. HEURISTIC STRUCTURE SEARCHING}

Hill-climbing method is directly used to select model structure in [13], where a satisfactory result can be obtained. However, the drawbacks of the hill-climbing still exist, and may limit the final performance and the computational complexity is still high. Therefore, we propose the following optimized selection method for model structures to enhance the hill-climbing method.

1) To enhance the global search capability of the hill-climbing and to reduce the probability of falling into local optima, multiple initial points are set and searched in a parallel way.

2) To search the feasible domain effectively, a black-list is established to prevent duplicate searches.

3) Actually, an appropriate set of basis is the primary contributor if it can best represent the parameters. We propose to use OMP method to further optimize the basis set to reduce model parameters, which is an efficient way since it can select the most appropriate subset.

4) Since heuristic algorithms including the hill-climbing method totally iterate in an adaptive way, the numerical instability is likely to happen especially when searching the high order models. Therefore, we also introduce a ridge regression technique into our optimized selection method to enhance numerical stability.

\subsection*{A. ENHANCED HILL-CLIMBING METHOD (EHC)}

In the initialization stage, in order to enhance global search capability of the hill-climbing and, to some extent, to avoid its likely trapping into local optimal solutions, multiple initial points instead of one are set and represented as

$$p_i, i = 1, ..., K$$

where initial point $p_i \in \mathbb{R}^{N_p \times 1}$ is set randomly and distributes in different places in the solution domain, which can support parallel search.
Take the $i$th search path as an example and compare the fitness values of the current search point with its neighbors, the next searching point will be decided

$$
\begin{align*}
     \begin{cases}
       p_i(n) = p_i^\delta(n) & \text{if } f_i^\delta < J_i \\
       p_i(n) = p_i(n - 1) & \text{if } f_i^\delta \geq J_i
     \end{cases}
\end{align*}
$$

(12)

where $p_i(n)$ is the $n$th iteration in the $i$th path, $p_i^\delta(n) \in U_i^\delta(n)$ is one of $p_i(n)$’s neighbors. The corresponding cost functions are $J_i^\delta$ and $J_i$. The fitness value of the next point is $f_i^\delta = \min f_i^\delta$. After the iteration, the optimal one from the multiple solutions obtained in parallel is taken as the final solution.

Traditionally, duplicated searches tend to appear in heuristic algorithms when searching different point with its neighbors, resulting in a low searching efficiency. For example, $p_i^\delta(n - 1)$ is a neighbor of the point $p_i(n - 1)$, which is also a neighbor of the current point as $p_i^\delta(n - 1) \in U_i^\delta(n - 1)$ and $p_i^\delta(n - 1) \in U_i^\delta(n)$. Since $U_i^\delta(n - 1) \cap U_i^\delta(n) \neq \emptyset$, there are some points are likely to be searched again or for several times, making it quite inefficient in both computation and memory resources, especially in a wideband scenario where a GMP model with high orders is needed.

Therefore, a black-list is created where the poor searched points are stored. It can avoid duplicate searching and thus improve search efficiency. Meanwhile, it can effectively avoid getting trapped in the previous search area leading to local optima. Specifically, add $\{\cdot\}_\text{sort}$ to $\text{Blacklist}$ in each iteration, where $\{\cdot\}_\text{sort}$ represents $L\%$ poor individuals in the neighborhood, and $J_{\text{temp}}$ is $J^\delta(n)$ sorted in ascending order.

In addition, accepting a suboptimal solution can also reduce the probability of being trapped into local optima, which is inspired by the simulated annealing (SA) [14], where

$$
p_i(n) = p_i^\delta(n) \text{ if } \text{random} > \text{Threshold} \tag{13}
$$

The flowchart of EHC algorithm is shown as follows:

\textbf{Algorithm. Enhanced Hill Climbing (EHC) algorithm}

\textbf{Input:} $P$ (Searching Paths)
\begin{itemize}
  \item $\delta$ (Radius of Neighborhood)
  \item $\beta$ (Trade-Off Factor)
\end{itemize}

\textbf{Output:} $p_{\text{best}}$ (Best points)

\begin{algorithmic}[1]
   \STATE for ($i = 1 : P$)
   \STATE Generate start point $p_i$ randomly, shown as (11);
   \FOR{$n = 1 : \text{MAX\_iteration}$}
   \STATE if $p_i^\delta(n) \in \text{BlackList}$
   \STATE choose $p_i^\delta(n) = \arg\min J^\delta_i(n)$;
   \STATE if $f_i^\delta(n) < J_i(n - 1)$:
   \STATE \quad $n = n + 1$;
   \STATE \quad $p_i(n) = p_i^\delta(n)$;
   \STATE \ELSE if random > Threshold:
   \STATE \quad $n = n + 1$;
   \STATE \quad $p_i(n) = p_i^\delta(n)$;
   \STATE \ENDIF
   \ENDIF
   \STATE if $n \sim = \text{MAX\_iteration}$
   \STATE $J_{\text{temp}} = \text{sort}(J^\delta_i(n), '\text{ascend}')$;
   \STATE $\text{temp} = [\text{Blacklist}_i J_{\text{temp}}]$;
   \STATE \ENDIF
   \ENDIF
   \ENDFOR
\end{algorithmic}

\textbf{B. PROPER SUBSET SELECTION OF BASIS FUNCTIONS}

In this section, we propose a method to further reduce model parameters. Actually, the performance of the model is contributed by an appropriate subset of model parameters, which means the possibility of using fewer basis functions in modeling. There are many approaches to implement further model size reduction.

Typically, principal component analysis (PCA) [24] is a well-known technique suitable for converting correlated bases into uncorrelated ones and reduce number of basis functions. Besides, orthogonal matching pursuit (OMP) [19] selects the primary proper subset by testing all the elements in the dictionary and sorting the contributions in reducing the NMSE. However, in DPD, OMP tends to achieve better performance than PCA while keeping the same number of basis functions. The reason is that after PCA, there tends to be more basis functions corresponding to certain components because of the decoupling of basis functions. Therefore, OMP is considered in our method.

\renewcommand{\arraystretch}{1.5}
\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
FLOPs & Corresponding to Different Operations \tabularnewline
\hline
\hline
Typically, principal component analysis (PCA) & A well-known technique suitable for converting correlated bases into uncorrelated ones and reducing the number of basis functions. Besides, orthogonal matching pursuit (OMP) selects the primary proper subset by testing all the elements in the dictionary and sorting the contributions in reducing the NMSE. However, in DPD, OMP tends to achieve better performance than PCA while keeping the same number of basis functions. The reason is that after PCA, there tends to be more basis functions corresponding to certain components because of the decoupling of basis functions. Therefore, OMP is considered in our method.
\end{tabular}
\end{table}
\renewcommand{\arraystretch}{1.0}


The computational complexity can be represented by:

\[ O^\Theta = 4K^2N + K^3 + 4K^2 + 4KN \]  \hspace{1cm} (14)

\[ O^@ = 4K^2N + 4KN + K^3 + 2K^2 - 4K \]  \hspace{1cm} (15)

where \( O^\Theta \) and \( O^@ \) represent the computational quantity of multiplications and additions, respectively, for one iteration.

### C. Numerical Stability Improvement

Using LS to calculate the coefficients containing matrix inversion operations tends to encounter larger error when the Hessian matrix involved is ill-conditioned hence leading to wrong iteration direction. Because the HC algorithm works in a fully adaptive manner, numerical instability is likely to happen especially when searching the high order models. Tikhonov regularization addresses the numerical instability of the matrix inversion and subsequently produces lower variance models. This method adds a positive constant to the diagonals of \( \mathbf{X}^T\mathbf{X} \), to ensure the matrix to be nonsingular.

As a consequence, the analytical expression can be represented as

\[ \min_{\alpha} \| \mathbf{x} \alpha - y \|^2 + \lambda \| \alpha \|^2 \]  \hspace{1cm} (16)

which is an unconstrained continuously differentiable convex optimization problem, with the solution in the form of

\[ \hat{\alpha}^{RR} = (\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^Ty \]  \hspace{1cm} (17)

where \( \lambda \) is the damping factor, \( \mathbf{I} \in \mathbb{C}^{K \times K} \) is an identity matrix. The method is called Ridge regression (RR) in [25], which is a particular case of Tikhonov regularization. The solution can be attained analytically in a closed-form.

A typical value of \( \lambda \) is equal to the resolution of the system. In this paper, we let \( \lambda = 2 \times 10^{-7} \). This choice is reasonable because a recent method of iterative updating \( \lambda \) is proposed in [26], and it converges to \( \lambda = 10^{-7} \sim 10^{-8} \), which is comparable to ours. This small bias does not disturb the solution evidently but significantly eases the ill-conditioning problem of the Hessian matrix because

\[ \kappa' = \frac{r_{\text{max}}'}{r_{\text{min}}'} = \frac{r_{\text{max}} + \sigma_1}{r_{\text{min}} + \sigma_2} \approx \frac{r_{\text{max}}}{\lambda_2} \ll \frac{r_{\text{max}}}{r_{\text{min}}} = \kappa \]  \hspace{1cm} (18)

where \( \kappa' \) is the matrix condition after modification, \( r_{\text{max}} \) and \( r_{\text{min}} \) are the maximum and the minimum eigenvalue of the Hessian matrix, \( \kappa \) is the original matrix condition. A larger condition number signifies an unstable numerical procedure calculating the matrix inversion. Therefore, the numerical stability improved significantly to avoid the instability that may lead to a wrong search direction.

We just take a simple model to show the important influence of numerical stability. As shown in Table II, final NMSE performance, matrix condition and floating-point operations (FLOPs) count of the Hessian Matrix involved in the single iteration are associated with model structures. The MP model is employed. An exhausting search method is adopted to determine the best structure, where the search domain is limited to memory depths 0 to 8 and nonlinear orders 1 to 8 including \( 9 \times 8 = 72 \) different possible models. The best model structure is \((\text{Mem.}, \text{Nonlin.}) = (4, 6)\).

Normally, with the increase of model orders, memory depths, or nonlinear orders, the matrix conditions and FLOPs will increase correspondingly as shown in Table II. Specifically, \( A(4,1) \) and \( B(4,6) \) have the same memory depth, and point B with higher nonlinear order has larger matrix condition numbers, more FLOPs, and lower NMSE. Similarly, A and C indicate that, under the same nonlinear order, increasing the memory depth will increase the matrix condition numbers and FLOPs. B represents the best MP structure. \( D(4,7) \) has the same memory depth and the nonlinear order is higher by 1 than B, but B has a better NMSE. Furthermore, the matrix condition number of D is nearly 1 212 times of B, and 2 001 560 more FLOPs are required. Therefore, choosing a suitable model size helps to achieve higher performance under lower matrix conditions and FLOPs.

### Table I: Computation Complexity of Each Operation

| Operation | Mult. | Add. |
|-----------|-------|------|
| \( \mathbf{X}^T\mathbf{X} \) | \( 4K^2N \) | \( 4K^2N - 2K^2 \) |
| \( (\cdot)^{-1} \) | \( K^3 \) | \( K^3 \) |
| \( \mathbf{X}^T\mathbf{y} \) | \( 4KN \) | \( 4KN - 2K \) |
| \( (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y}) \) | \( 4K^2 \) | \( 4K^2 - 2K \) |

So far, in order to solve the problems existing in traditional HC algorithm, including likely falling into the local optimal solution affecting the final performance, often duplicated the search resulting in very low efficiency, and unstable numerical stability during the iteration process, we propose EHC, which has more powerful global search ability through multiple initial points parallel search and borrowed idea from SA. And the blacklist in EHC can effectively prevent duplicate searches and improve search efficiency. Moreover, RR is used to enhance the numerical stability, which makes EHC search more robust.
Next, in order to illustrate the global search ability of the proposed algorithm, we further propose EGA algorithm with global search capability for more comparison.

Genetic algorithm (GA) is also an important member of the heuristic algorithms, which is a classical global optimal adaptive search algorithm. Therefore, it is chosen as a comparison of the global search capabilities of EHC. In addition, we also use RR in GA to enhance numerical stability and introduce a blacklist mechanism to prevent duplicated searches. We call it enhanced genetic algorithm (EGA).

In EGA, the fitness function can be defined in (10). Crossover and mutation operators are used to search for feasible regions. A crossover sample is presented in Fig. 4. For \((P_1^i, P_2^i)\), interchanges happen in the intersections to produce new individuals \((P_1^i', P_2^i')\). Then, two individuals with best fitness are selected in \((P_1^i, P_2^i, P_1^i', P_2^i')\) as the next generation \((P_1^{i+1}, P_2^{i+1})\) to participate next iteration. Mutation operator refers to the process of replacing certain genetic values in a chromosome with other values to create a new one. A mutation example is also presented in Fig.4, where gray depth represents the changes in the corresponding index of the origins limited by the search domain.

The crossover and mutation operations cooperate with each other to search the model structure in the search domain. A concise experimental result is presented in TABLE III to confirm the feasibility and effectiveness of the EGA. The specific parameter selection process for EGA is in the next section.

So far, we propose two search methods based on heuristic algorithm, namely EHC and EGA. In the subsequent experiments, on the one hand, we specifically verify the high performance of models obtained by EHC and EGA. On the other hand, we also compare the global search capability of EHC with EGA to verify that the proposed EHC method has comparable global search capability with EGA.

### TABLE III

| Method | Num. of Coeff. | NMSE (dB) | ACPRL/AC-PRU (dBc) |
|--------|---------------|-----------|---------------------|
| w/o DPD | -             | -20.02    | -32.17/-33.45       |
| GA     | 40            | -40.01    | -49.30/-49.35       |

EHC with EGA to verify that the proposed EHC method has comparable global search capability with EGA.

### E. PROPOSED ARCHITECTURE

In order to combine the proposed model search algorithms with DPD in a more flexible way. A novel structure is proposed and the block diagram is shown in Fig. 3. Firstly, the heuristic algorithm proposed above is used offline to search a model structure. Then the extracted model structure is employed in DPD. \(\tilde{x}(n), \tilde{y}(n), \tilde{z}(n) \in \mathbb{C}^{N \times 1}\) represent complex signal.
input signal, output signal, and the signal after DPD, respectively. \( p_{\text{best}} \) is the model structure searched by heuristic algorithm we proposed above. The proposed architecture consists of the following three working modes.

| Mode       | \( SW - I \) | \( SW - II \) |
|------------|--------------|---------------|
| Initialization | On        | Off          |
| Working     | Off         | On           |
| Updating    | On          | On           |

The working mode of the proposed architecture is shown in TABLE IV including:

- Initialization) \( SW - I \) is on and \( SW - II \) is off. Either heuristic method proposed above is employed to obtain an optimal model marked as \( p_{\text{best}} \) in an offline way.
- Working) \( SW - I \) is off and \( SW - II \) is on. DPD training is carried out under IDL structure based on the optimal model \( p_{\text{best}} \) obtained in initialization mode or updating mode.
- Updating) Both \( SW - I \) and \( SW - II \) are on. Once the operating state of the PA changes significantly, the DPD model should be updated. The above heuristic method operates again to obtain a new optimization model \( p_{\text{best}} \) offline without interrupting the operation of DPD. It should be noted that the updating mode operates offline and the parallel mechanism introduced can accelerate the convergence.

### IV. EXPERIMENTAL VALIDATION

#### A. MEASUREMENT SETUP

In this section, the proposed model is validated and compared with the experimental model. The photograph of the experimental platform is shown in Fig. 5, which is composed of a personal computer (PC) with MATLAB R2021a, a Vector Signal Generator (SMW200A) from Rohde & Schwarz, a linear drive amplifier, the main PA to be tested (BLM9D2325-20AB), a spectrum analyzer, a 40dB RF attenuator.

The specific procedure is as follows: firstly, the discrete baseband signal generated by PC is sent to the Vector Signal Generator where a continuous baseband OFDM signal is generated and up-converted to the RF band. Secondly, the RF signal is amplified by the linear PA and the main PA. Thirdly, the output of the main amplifier is attenuated, down-converted, and sent to the spectrum analyzer. Finally, the signal is sent back to the PC for DPD processing, specifically including power-alignment, time-alignment, and coefficients update.

#### B. HEURISTIC PARAMETERS SETTING

Before the experiment, parameters of the heuristic need to be determined. The signal used to determine heuristic parameters is a 100 MHz-OFDM and the tested PA mentioned above. The range of the 8 dimensional parameters in (1) varies according to different application scenarios. The regularization factor \( \beta \) in (10) is searched offline at an interval of 0.0001 between 0.001 and 0.2, and a best value of 0.036 is obtained.
In GMP model. The comparison model will be marked as memory depth, we need to ensure the existence of cross terms. Taken to be close but not greater than leading and lagging terms. In terms of nonlinearity, capability but also achieve more efficient search.

Of EHC is still close to the convex hull even with the enhanced search domain which contains 3732480 different possible models. The model structure used for comparison is obtained, following the “Comparison Model Setting” as mentioned above, to be \( P_{\text{compa}} = [5, 10, 5, 2, 2, 5, 2, 2] \) specifically.

In the initialization mode, turn on \( SW - I \) in Fig. 3 and obtain \( P_{\text{best}} \) based on the proposed heuristic algorithms. In EHC, the searched best model is \( P_{\text{EHC}} = [6, 2, 5, 1, 4, 1] \), and EGA converges to model \( P_{\text{EGA}} = [4, 5, 2, 3, 2, 2, 2] \). In terms of the highest memory depth and non-linearity, there is no much difference between the final model obtained by EHC and EGA, specifically. In terms of the number of parameters, EHC has 48 parameters and EGA has 44 parameters. And the performance of the two models is comparable.

In order to further illustrate, a neighborhood of radius \( \delta = 2 \) of \( P_{\text{EHC}} \) and \( P_{\text{EGA}} \) are presented in Fig. 7 marked as two green diamonds. Both \( P_{\text{EHC}} \) and \( P_{\text{EGA}} \) are the structures with the best performance in their respective neighborhoods. These two models obtained are then further simplified by OMP represented by two black stars, where NMSE are comparable and superior to other DPD structures with the same number of parameters.

Then turn off \( SW - I \), and turn on \( SW - II \). The model \( P_{\text{best}} \) is trained based on IDL to obtain the corresponding DPD coefficients. The convergence curves of different models determined by EHC and EGA model are shown in Fig. 8. There is almost no difference in convergence speed between them. The specific NMSE, ACPR/ACPRU and EVM are listed in TABLE VI, and the proposed method can find models with comparable performance and less parameters. Obviously, for the case without DPD, because the PA operated in a high gain compression state, a poor NMSE is got. For the other four cases with DPD, comparable and satisfying performance is got. After the model converges, the power spectral density (PSD) comparison among different models is shown in Fig. 9. Compared with the case without DPD all models have comparable and quite satisfactory adjacent channel performance after DPD.

Finally, compared with the Compara model, the computational complexity of models searched by EHC and EGA are both reduced by approximately 75%, which is calculated based on (14) and (15) and the results are shown in 60 MHz section in Fig. 10. However, the searched models of EHC and EGA exist difference. In EHC, several search paths instead of one can improve search capabilities and avoid being trapped into local optima.

| TABLE V | HEURISTIC PARAMETERS SETTING THROUGH EXPERIMENTS |
|---------|-----------------------------------------------|
|         | EHC | EGA |
| P       | 3   | 5   | 80  | 100 | 120 |
| Opt.    | -40.81 | -40.79 | -37.75 | -40.71 | -40.73 |
| NMSE\_db | 7.60 | 9.13 | 5.08 | 7.74 | 9.01 |

Under the above settings, Fig. 6 can be obtained. The discrete red dots are structures searched by EGA, and the blue ones represent structures searched by EHC.

The convex hull of the search interval formed by the red stars is also called the Pareto front, which is composed of the optimal solution set. The convex hull represents the best trade-off between model performance and complexity. In fact, the problem of finding suitable model structure is a multi-objective optimization problem, which aims to find the best balance between model complexity and performance. In such problems, the optimal solution set of the objective function also becomes the Pareto optimal solution set.

Furthermore, from the distribution in Fig. 6, we can confirm that the global search ability of EHC with multiple initial points parallel search is enhanced greatly. Because of the strong local search capability of EHC itself, the search range of EHC is still close to the convex hull even with the enhanced global search capability. Therefore, each path in EHC can converge to the convex hull quickly. It can be concluded that compared with EGA, EHC can not only enhance the global search capability but also achieve more efficient search.

C. COMPARISON MODEL SETTING

For a fairer comparison, we select the GMP comparison model according to the following methodology.

Since GMP model can be considered as an extended version of MP model, hence to determine the parameters of the comparison GMP model, we first select the best memory depth and nonlinear order in MP model as mentioned in part C of section II, and use them as \( M_a \) and \( K_a \) in GMP model, respectively. After that, determine the parameters of the leading and lagging terms. In terms of nonlinearity, \( K_b, K_c \) are taken to be close but not greater than \( K_a \). And in terms of memory depth, we need to ensure the existence of cross terms. It should be noted that in general, \( L_b \geq M_b \), and \( L_c \geq M_c \) are in GMP model. The comparison model will be marked as Compara in the following section.

D. MEASUREMENT RESULTS

Experiments on both 60 MHz and 100 MHz signal are conducted and they will be described in detail as follows.

1) 60 MHz SIGNAL

The test signal employed in the measurement is a 60 MHz OFDM signal with 8.74 dB peak to average power ratio (PAPR). The length of the data is 96 176. The main PA is driven under a high compression with -4 dBm input power. The sampling rate is 368.64 MHz. And that \( 1 \leq K_a \leq 8, 1 \leq L_a \leq 10, 0 \leq K_b \leq 5, 0 \leq L_b \leq 5, 0 \leq M_b \leq 5, 0 \leq K_c \leq 5, 0 \leq L_c \leq 5 \) and \( 0 \leq M_c \leq 5 \) are used to specify the search domain which contains 3 732 480 different possible models. The model structure used for comparison is obtained, following the “Comparison Model Setting” as mentioned above, to be \( P_{\text{compa}} = [5, 10, 5, 2, 2, 5, 2, 2] \) specifically.

In the initialization mode, turn on \( SW - I \) in Fig. 3 and obtain \( P_{\text{best}} \) based on the proposed heuristic algorithms. In EHC, the searched best model is \( P_{\text{EHC}} = [6, 2, 5, 1, 4, 1] \), and EGA converges to model \( P_{\text{EGA}} = [4, 5, 2, 3, 2, 2, 2] \). In terms of the highest memory depth and non-linearity, there is no much difference between the final model obtained by EHC and EGA, specifically. In terms of the number of parameters, EHC has 48 parameters and EGA has 44 parameters. And the performance of the two models is comparable.

In order to further illustrate, a neighborhood of radius \( \delta = 2 \) of \( P_{\text{EHC}} \) and \( P_{\text{EGA}} \) are presented in Fig. 7 marked as two green diamonds. Both \( P_{\text{EHC}} \) and \( P_{\text{EGA}} \) are the structures with the best performance in their respective neighborhoods. These two models obtained are then further simplified by OMP represented by two black stars, where NMSE are comparable and superior to other DPD structures with the same number of parameters.

Then turn off \( SW - I \), and turn on \( SW - II \). The model \( P_{\text{best}} \) is trained based on IDL to obtain the corresponding DPD coefficients. The convergence curves of different models determined by EHC and EGA model are shown in Fig. 8. There is almost no difference in convergence speed between them. The specific NMSE, ACPR/ACPRU and EVM are listed in TABLE VI, and the proposed method can find models with comparable performance and less parameters. Obviously, for the case without DPD, because the PA operated in a high gain compression state, a poor NMSE is got. For the other four cases with DPD, comparable and satisfying performance is got. After the model converges, the power spectral density (PSD) comparison among different models is shown in Fig. 9. Compared with the case without DPD all models have comparable and quite satisfactory adjacent channel performance after DPD.

Finally, compared with the Compara model, the computational complexity of models searched by EHC and EGA are both reduced by approximately 75%, which is calculated based on (14) and (15) and the results are shown in 60 MHz section in Fig. 10. However, the searched models of EHC and EGA exist difference. In EHC, several search paths instead of one can improve search capabilities and avoid being trapped into local optima.

| TABLE VI | A PERFORMANCE PRESENTATION OF THE MODEL SEARCHED USING THE PROPOSED METHOD WITH 60 MHZ INPUT SIGNALS |
|-----------|-------------------------------------------------|
| Method    | Coeff. Num. | NMSE (dB) | ACPR/ACPRU (dBc) | EVM (%) |
| w/o DPD   | -            | -21.65    | -31.11/-33.50    | 7.01    |
| Compara.  | 79           | -40.52    | -48.81/-49.05    | 0.59    |
| EHC       | 48           | -40.53    | -49.10/-49.29    | 0.59    |
| EHC+OMP   | 39           | -40.80    | -49.23/-49.25    | 0.56    |
| EGA       | 44           | -41.13    | -49.14/-49.21    | 0.51    |
| EGA+OMP   | 40           | -41.01    | -49.30/-49.35    | 0.54    |
While in EGA, a relatively optimal result is reached because of its global search capability.

2) 100 MHz SIGNAL
The test signal is an OFDM signal with an increased bandwidth of 100 MHz, and its PAPR is 8.78 dB. The length of data is 92 112. The input power is -3 dBm and the sampling rate is 614.4 MHz. The main PA is a Doherty PA from Ampleon with a center frequency at 2.4 GHz.

Compared with case 1), the bandwidth of the input signal is increased, implying stronger nonlinearity and memory effect of the PA. The order of the model should be increased accordingly, in order to improve the modeling ability. Ranges of the coefficients are set as $1 \leq K_a \leq 8$, $1 \leq L_a \leq 20$, $0 \leq K_b \leq 5$, $0 \leq L_b \leq 5$, $0 \leq M_b \leq 8$, $0 \leq K_c \leq 8$, $0 \leq L_c \leq 8$, and $0 \leq M_c \leq 8$ to limit the search domain containing 7 464 960 different possible model structures. And according to the “Comparison Model Setting” mentioned above, the comparison model is obtained as $P_{\text{compara}} = [6, 12, 5, 4, 4, 4, 4, 4]$.

Turn on $SW - I$ in Fig. 3, and obtain $P_{\text{test}}$. In EHC, the final model $P_{\text{EHC100M}} = [7, 4, 5, 1, 1, 3, 5, 3]$ is obtained and 909 different models are compared. The EGA converges to $P_{\text{EGA100M}} = [7, 4, 3, 4, 0, 4, 5, 2]$ with 808 different models compared. It can be seen that the models searched by the two methods are very similar.

Although there are slightly differences between their results, they all achieve a good trade-off between complexity and performance. After OMP, the number of parameters in EGA and EHC is reduced to 66 and 68, respectively. In Fig. 11, we compare all models with a neighborhood radius $\delta = 2$ for $P_{\text{EHC100M}}$ and $P_{\text{EGA100M}}$ that are the structures with the best performance in their respective neighborhoods. After OMP, parameters number of $P_{\text{EHC100M}}$ and $P_{\text{EGA100M}}$ are further reduced, and they significantly outperform other structures with the same number of parameters. The AMAM and AMPM of PA are shown in Fig.12, which are calculated and plotted offline, and it can be seen that the PA has been effectively linearized.

| Table VII | A PERFORMANCE PRESENTATION OF THE MODEL SEARCHED USING THE PROPOSED METHOD WITH 100 MHz INPUT SIGNALS |
|------------|--------------------------------------------------------------------------------------------------|
| Method     | Coeff. Num. | NMSE (dB)   | ACPRL/ACP-PRU (dBc) | EVM (%) |
| w/o DPD    | -           | -20.02      | -32.17/-33.45       | 9.21    |
| Compar.    | 222         | -40.82      | -48.68/-48.17       | 0.72    |
| EHC        | 83          | -40.23      | -48.67/-48.10       | 0.73    |
| EHC+OMP    | 68          | -40.53      | -48.68/-48.13       | 0.78    |
| EGA        | 75          | -40.70      | -48.80/-48.28       | 0.72    |
| EGA+OMP    | 66          | -40.93      | -48.71/-48.61       | 0.76    |
FIGURE 11. Neighborhood of radius $\delta = 2$ of $P_{EHC_{100}}$ and $P_{EGA_{100}}$, where 238 models are included in the neighborhood of $P_{EHC_{100}}$ and 211 models are included in the neighborhood of $P_{EGA_{100}}$.

FIGURE 12. AMAM and AMPM of the PA with 100 MHz input signal, (a). AMAM comparison. (b) AMPM comparison.

Then, turn off $SW-I$ and turn on $SW-II$ is closed for DPD training. The optimal solution and their convergence rate are comparable with each other as shown in Fig. 13. The specific numerical results are listed in TABLE VII, where the NMSE, ACPRU/ACPRL are comparable with the $P_{compa}$ as shown in Fig. 13 and Fig. 14, and the EVM is comparable too. Besides, the models generated by heuristic proposed have less parameters, which greatly reduces the computational complexity. For both EGA and EHC, the amount of computation in a single iteration is reduced by approximately 80% as shown in the 100 MHz section in Fig. 10.

V. CONCLUSIONS

In this paper, a novel method based on the heuristic algorithm is proposed to optimize and identify the appropriate DPD model structure with two specific algorithms, namely EHC and EGA. The proposed method explores the problem of obtaining an optimal solution in a considerable large search domain in parallel, and this search domain is difficult to search exhaustively. Once the operating state of the PA changes significantly, new optimal structures can be searched offline also in parallel with DPD.

FIGURE 13. Convergence curves of different models with 100MHz input signal.

FIGURE 14. PSD of different models with 100 MHz input signal.

Then, since the heuristic algorithm itself iterates in an adaptive manner, RR is introduced to improve the numerical stability and ensure the correctness of the search direction. Furthermore, considering the fact that only a proper set closely related to the parameters is actually the primary contributor to the final performance, OMP is introduced to further reduce the number of parameters. Therefore, the combination of parallel searching, enhancing the numerical stability with RR and further reduction of model parameters using OMP greatly reduces the computational complexity and improves the robustness of the searched model. Specifically, the proposed EHC has a more powerful global search ability and can effectively prevent duplicated searches and improve search efficiency. And both EHC and EGA have a more stable and faster iterative process owning to the introduction of RR and blacklist.

Finally, based on the novel heuristic DPD paradigm we proposed, the experimental results of GMP on both 60 MHz and up to 100 MHz bandwidth signals further prove the effectiveness of the proposed method.

ACKNOWLEDGMENT

This research is partially supported by the National Natural Science Foundation of China under Grant No. 61471333. We
also thank the Information Science Center of USTC for experimental support.

REFERENCES

[1] S. Muntaz, J. Rodrigues, L. Dai, *muWave Massive MIMO: A Paradigm for 5G*, New York, NY, USA: Academic, 2016.

[2] G. Li, Y. Zhang, H. Li, W. Qiao, and F. Liu, “Instant gated recurrent neural network behavioral model for digital predistortion of rf power amplifiers,” IEEE Access, vol. 8, pp. 67 474–67 483, Apr. 2020.

[3] H. Li, W. Qiao, Q. Liu, G. Li, Y. Zhang, G. Wang and F. Liu, “Signed orthogonal regressor algorithm for digital predistortion of power amplifiers,” IEEE Microw. Wireless Compon. Lett., vol. 31, pp. 809-872, July 2021.

[4] N. Guan, N. Wang, H. Wang, “Digital predistortion of wideband power amplifier with single undersampling ADC,” IEEE Microw. Wireless Compon. Lett., vol. 27, no. 11, pp.1016-1018, Nov. 2017.

[5] F. M. Ghannouchi and O. Hammi, “Behavioral modeling and predistortion,” IEEE Microw. Mag., vol. 10, no. 7, pp. 52-64, Dec. 2009.

[6] L. Guan and A. Zhu, “Green communications: digital predistortion for wideband rf power amplifiers,” IEEE Microw. Mag vol. 15, no. 7, pp. 84-99, Nov.-Dec. 2014.

[7] S. Boyd, L. O. Chua, and C. A. Desoer. “Analytical foundations of volterra series.” *IMA J Math Control Info* 1.3(1984):243-282.

[8] L. Ding et al., “Memory polynomial predistorter based on the indirect learning architecture,” in *IEEE Globecom Conf.,* Nov. 2002.

[9] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt and J. Pastalan, “A generalized memory polynomial model for digital predistortion of rf power amplifiers,” IEEE Trans. Signal Process., vol. 54, no. 10, pp. 3852-3860, Oct. 2006.

[10] L. Guan and A. Zhu, “Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers,” in *Proc. IEEE Int. Integ. Nonlinear Microw. Millimeter-Wave Circuits Workshop*, Vienna, Austria, Apr. 2011, pp. 1-4.

[11] A. Tehrani, H. Cao, S. Afsardoust, T. Eriksson, M. Isaksson, and C. Fager, “A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models,” IEEE Trans. Microw. Theory Techn., vol. 58, no. 6, pp. 1510–1520, June 2010.

[12] S. Wang, M. A. Hussein, O. Venard and G. Baudoin, “Optimal sizing of generalized memory polynomial model structure based on Hill-Climbing heuristic,” in *2016 46th European Microw. Conf. (EuMC)*, 2016, pp. 190-193.

[13] R. Mondal, T. Ristaniemi and M. Doula, “Genetic algorithm optimized memory polynomial digital pre-distorter for rf power amplifiers,” in *2013 Int. Conf. on Wireless Communications and Signal Processing*, 2013, pp. 1-5.

[14] Hwang, C. R., “Simulated annealing: Theory and applications.” *Acta Applicandae Mathematicae* 12.1(1988):108-111.

[15] A. H. Abdelhalafiz, O. Hammi, A. Zerguine, A. T. Al-Awami and F. M. Ghannouchi, “A pso based memory polynomial predistorter with embedded dimension estimation,” *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 665-673, Dec. 2013.

[16] O. Hammi, A. M. Kedir and F. M. Ghannouchi, "Nonuniform memory polynomial behavioral model for wireless transmitters and power amplifiers,” in *2012 Asia Pacific Microw. Conf. Proceedings*, Kaohsiung, Taiwan, 2012, pp. 836-838.

[17] H. Yu, G. Xu, T. Liu, J. Huang and X. Zhang, "A memory term reduction approach for digital pre-distortion using the attention mechanism," *IEEE Access*, vol. 7, pp. 38185-38194, March 2019.

[18] H. Wang, J. Bao and Z. Wu, “Comparison of the behavioral models for rf power amplifier with memory effects,” IEEE Microw. Wireless Compon. Lett., vol. 19, no. 3, pp. 179-181, March 2009.

[19] J. A. Becerra, M. J. Madero-Ayora and C. Crespo-Cadenas, “Comparative analysis of greedy pursuits for the order reduction of wideband digital predistorters,” IEEE Trans. Microw. Theory Techn., vol. 67, no. 9, pp. 3575-3585, Sept. 2019.

[20] A. Zhu, Draxler, P.J., Yan, J.J., Brazil, T.J., Kimball, D.F., and Asbeck, P.M., “Open-loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series,” IEEE Trans. Microw. Theory Techn., vol. 56, no. 7, pp. 1524-1534, June 2008.

[21] N. Guan, N. Wang, H. Wang, “Model identification for digital pre-distortion of power amplifier with signed regressor algorithm,” IEEE Microw. Wireless Compon. Lett., vol.28, no. 10, pp. 921-923, Oct. 2018.

[22] O. Hammi, F. M. Ghannouchi, and B. Vassilakis, “A compact envelope-memory polynomial for rf transmitters modeling with application to baseband and RF-digital predistortion,” IEEE Microw. Wireless Compon. Lett., vol. 18, no. 5, pp. 359-361, May 2008.

[23] M. Schetzen, “Theory of p-th-order inverses of nonlinear systems,” IEEE Trans. Circuits Syst., vol. 23, no. 5, pp. 285-291, May 1976.

[24] D. Lopez–Bueno, Q. A. Pham, G. Montoro and P. L. Gilabert, “Independent digital predistortion parameters estimation using adaptive principal component analysis,” IEEE Trans. Microw. Theory Techn., vol. 66, no. 12, pp. 5771-5779, Dec. 2018.

[25] L. Guan, and A. Zhu, “Optimized low-complexity implementation of least squares based model extraction for digital predistortion of RF power amplifiers,” IEEE Trans. Microw. Theory Techn., vol. 60, no. 3, pp. 594-603, Mar. 2012.

[26] Y. Liu, X. Xia, Q. Xu, W. Pan, and Y. Tang, “Relaxing requirements on training samples in digital predistortion by using ridge regression.” IEEE Microw. Wireless Compon. Lett., PP.99(2021):1-1.