Computational and Structural Biotechnology Journal 23 (2024) 2837–2850

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

Exploring sperm cell motion dynamics: Insights from genetic algorithm-based analysis

Anke Klingner [a], Alexander Kovalenko [b], Veronika Magdanz [c], Islam S.M. Khalil [d,*]

[a] Department of Physics, German University in Cairo, New Cairo, 11835, Egypt
[b] Faculty of Information Technology, Czech Technical University in Prague, Prague, 16000, Czech Republic
[c] Department of Systems Design Engineering, Waterloo Institute for Nanotechnology, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[d] Department of Biomechanical Engineering, University of Twente, Twente, 7500 AE, the Netherlands

ARTICLE INFO

Keywords:
Sperm cell dynamics
Genetic algorithm
Flagellum deformation
Motion analysis
Biological motion

ABSTRACT

Accurate analysis of sperm cell flagellar dynamics plays a crucial role in understanding sperm motility as flagella parameters determine cell behavior in the spatiotemporal domain. In this study, we introduce a novel approach by harnessing Genetic Algorithms (GA) to analyze sperm flagellar motion characteristics and compare the results with the traditional decomposition method based on Fourier analysis. Our analysis focuses on extracting key parameters of the equation approximating flagellar shape, including beating period time, bending amplitude, mean curvature, and wavelength. Additionally, we delve into the extraction of phase constants and initial swimming directions, vital for the comprehensive study of sperm cell pairs and bundling phenomena. One significant advantage of GA over Fourier analysis is its ability to integrate sperm cell motion data, enabling a more comprehensive analysis. In contrast, Fourier analysis neglects sperm cell motion by transitioning to a sperm-centered coordinate system (material system). In our comparative study, GA consistently outperform the Fourier analysis-based method, yielding a remarkable reduction in fitting error of up to 70% and on average by 45%. An in-depth exploration of the sperm cell motion becomes indispensable in a wide range of applications from complexities of reproductive biology and medicine, to developing soft flagellated microrobots.

1. Introduction

Flagella are whip-like appendages ubiquitously found in a diverse array of organisms, which range from single-celled bacteria such as *Escherichia coli* and *Salmonella* or archaea, and extend to eukaryotic species inclusive of certain algae and protozoans like *Giardia lamblia*. Interestingly, in the realm of multicellular organisms, flagella adorn the male gametes, functioning as pivotal facilitators of motility. The physical aspects of flagellar beating have been comprehensively summarized by Holwill [11].

The flagellum, primarily drives an organism's propulsion, thereby aiding in a multitude of vital activities such as nutrient acquisition, predator evasion, and ensuring optimal environmental conditions for the organism's survival. However, the importance of studying flagella dynamics goes beyond the understanding of immediate locomotion mechanics [2,6,7,10,13].

A deep comprehension of flagellar dynamics can unveil detailed behaviors of microorganisms at a granular level, offering insights into their ecological roles and contributing to the development of strategies for managing pathogenic species [3]. In multicellular entities, where the male reproductive cell, or sperm cell, possesses a flagellum, thoroughly investigating this feature [9] is crucial for understanding the nuances of male reproductive health [16]. Such understanding can guide both the diagnosis and treatment of male fertility issues. Additionally, expertise in flagellar dynamics can enhance assisted reproductive methods. For example, in artificial insemination and *in-vitro* fertilization, an advanced grasp of flagellar movement might boost the effectiveness of these procedures [26]. Additionally, flagella serve as biological blueprints in the field of micro-robotics [12,14], inspiring the design of micro-robots capable of navigating fluidic environments with adept precision. This presents considerable implications for precision medicine, particularly in the realm of drug delivery [15,30], by paving the way for targeted treatment modalities that minimize side effects and maximize therapeutic efficacy. Therefore, comprehensive study of flagellar dynamics is not only key to deepening our understanding of biological systems but is also poised to drive advancements in various fields spanning medicine,

* Corresponding author.
E-mail address: i.s.m.khalil@utwente.nl (I.S.M. Khalil).

https://doi.org/10.1016/j.csbj.2024.06.008
Received 18 March 2024; Received in revised form 5 June 2024; Accepted 6 June 2024
Available online 27 June 2024

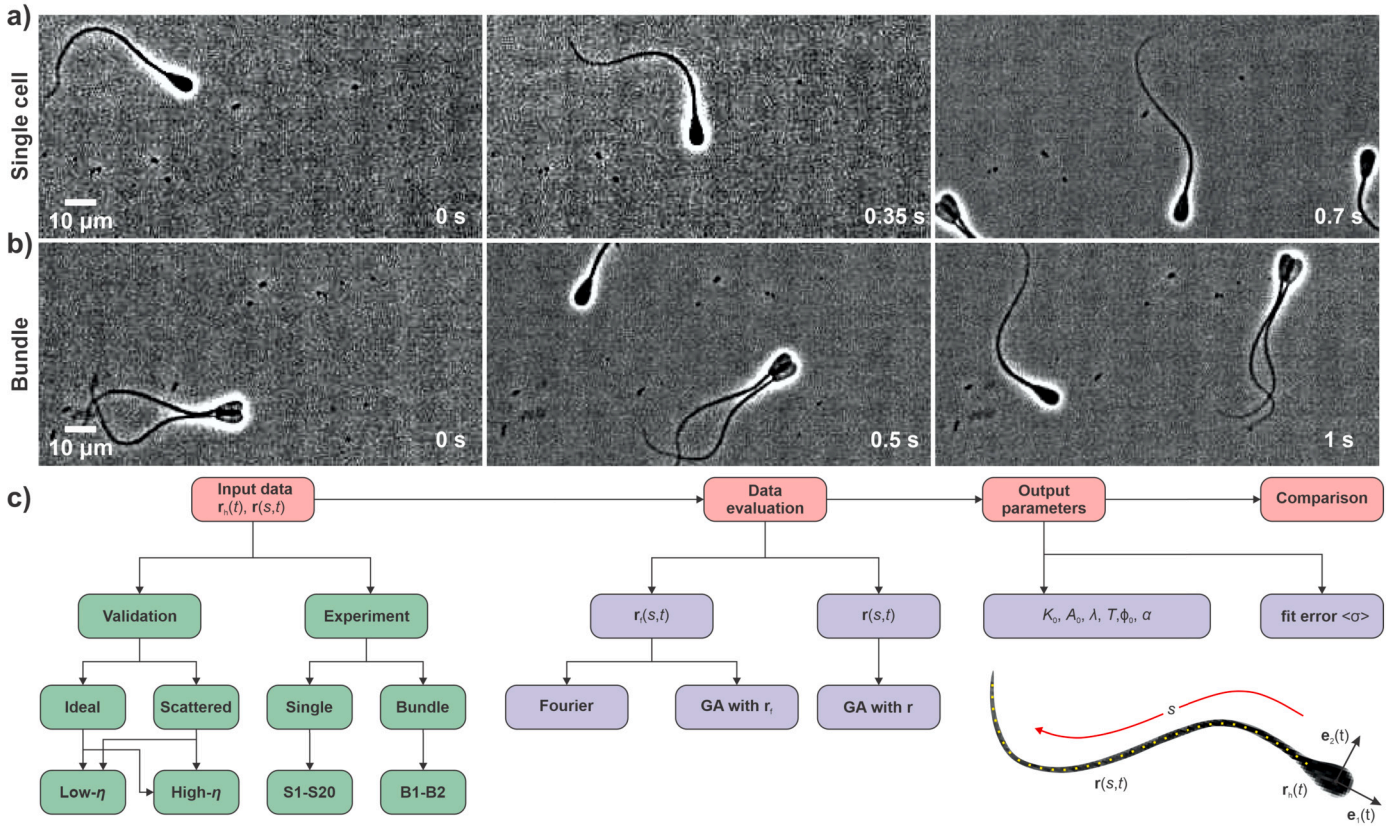

Fig. 1. Videos covering several beat cycles of a) single cells and b) bundles of two sperm cells are used to extract experimental input data such as temporal sperm head position $\mathbf{r_h}(t)$ and spaciotemporal flagellum shape $\mathbf{r}(s,t)$ in laboratory coordinate system. c) Schematic flowchart of implemented processes for input data acquisition, data evaluation methods, and output parameter extraction. Beside experimental input data, ideal and scattered validation data is created for testing of the data evaluation methods. Then, data evaluation methods Fourier, GA with $\mathbf{r_f}$ and GA with $\mathbf{r}$ are applied. They are either based on input data in laboratory system or on data $\mathbf{r_f}(s,t)$ which is the spaciotemporal flagellum shape in shifted material coordinate system fixed to position of sperm head. Output parameters mean curvature K_0, beating amplitude A_0, wavelength λ, period time T, initial phase shift ϕ_0, initial swimming direction α and fit error $<\sigma>$ are extracted and compared.

technology, and environmental science. In this regard using automated Genetic Algorithms (GA) [18,27,28] could be a step forward to obtain a deeper insight into the abovementioned problem.

GA are generic population-based metaheuristic [29] optimization algorithms, inspired by natural selection, that mimic processes like mutation, crossover, and selection to produce superior solutions for optimization challenges. In this method, a population of candidates evolves across generations. Each member's fitness is evaluated, and the most adept ones advance to subsequent generations. GA due to their explorative nature are suitable for global optimization problems and capable of exploring the entire solution space. GA can perform reasonably well, as the fitness landscape changes, the algorithm can still locate optimal or near-optimal solutions. Unlike neural networks, GA do not require the problem to be differentiable, they are able to adapt to changing environments or problem definitions over time and can work with any fitness function, whether it is well-defined or not. Moreover, GA are inherently parallel which can be leveraged by modern multicore hardware.

The objective of this study is to employ GA in analysis of single sperm cell and sperm bundle motion, as illustrated in Fig. 1. Notably, GA have the capability to extract a broader range of parameters from experiments compared to the commonly used Fourier analysis. The ensuing section delineates the implementation process, as summarized in Fig. 1. It elucidates the generation of input data, subsequent evaluation using various methodologies to derive output parameters. These resultant parameters are then juxtaposed across different techniques and input datasets. Ultimately, the study culminates in a comprehensive comparison and discussion of all obtained results.

2. Problem overview and related work

The propulsion of the male gamete in vertebrates is facilitated by a eukaryotic flagellum. Central to this structure is a 9+2 axoneme of microtubules, a feature ubiquitous across most cilia and flagella. Bull sperm cells have been a focal point of research due to their economic relevance. These cells, with an average length of about 60 µm, have a period time of nearly 30 ms and can attain speeds between 40 and 150 µm/s [8,11,20–23]. The oscillatory movement of the flagellum is analogous to a wave in motion. Subsequently, the crux of sperm cell analysis lies in identifying the salient parameters of this wave [11]. Key parameters of bull spermatozoa, the model system used in this study, encompass a wavelength of 33 to 62 µm, a beating amplitude between 6.7 and 8.5 µm, and a period time from 42 to 111 ms. The zeroth and first orders of Fourier decomposition have been demonstrated to encompass over 95% of flagellar power. Consequently, parameters such as wavelength, period time, amplitude, and mean curvature are adequate for describing flagellar beating [19]. Optimal velocities are recorded at temperatures of 38 °C [24]. Interestingly, with a viscosity range of 0.8 to 1.6 mPa.s, the velocity amplifies, but it diminishes at 10 mPas [5]. At heightened viscosities of around $2-3$ Pa.s, the flagellum showcases planar movements as opposed to the helical patterns at lower viscosities [25].

Various research efforts have quantified the energy expenditure of bull sperm cell, indicating values like 2.1×10^{-14} J/s/flagellum at 37 °C [25] and 2×10^{-13} J/s/flagellum [21]. Mechanical properties such as the Young's modulus for different components of bull spermatozoa have been evaluated [22]. Detailed observations reveal a nonsinusoidal wave

Algorithm 1 Genetic algorithm for flagellar dynamics.

Require: Generations $N_G = 1000$, population size $N_P = 20$
Require: Mutation and crossover prob., $p_m = 0.2$, $p_c = 0.1$
Chromosome Representation: Each chromosome is represented as: $[K_0, A_0, \lambda, T, \phi_0, \alpha; L, D, 2a, 2b, \eta]$
Initialize a population of N_P individuals
for each individual **do**
  Randomly set K_0 in the range $(-30, 30)$, A_0 in $(1, 30)$
  Randomly set λ in $(30, 100)$, T in $(20, 600)$
  Randomly set ϕ_0 in $(-180, 180)$, α in $(-180, 180)$
  Set L in $(60, 65)$, $D = 1$, $2a = 5$, $2b = 2.5$, η
end for
Initialize a chromosome using Fourier method results for one individual.
for generation in 1 to N_G **do**
  Evaluate the fitness $< \sigma >$ of each individual.
  Select two parents randomly.
  if random value $< p_c$ **then**
    Perform crossover on parents to produce offspring.
    if Fitness of offspring higher than parent **then**
      Replace parent with offspring.
    end if
  end if
  for each individual **do**
    for each gene **do**
      if random value $< p_m$ **then**
        Mutate gene within the allowed range.
      end if
    end for
    if Fitness of mutant higher than for individual **then**
      Replace individual with mutant
    end if
  end for
  Select best individual into new population
  for $0.75N_P$ **do**
    Choose the superior individual from two random selections for the new population
  end for
  Initialize individuals randomly to reach N_P individuals in new population
  Replace old population with new population.
end for
Select the best individual from the last generation.
return Best individual's chromosome as the solution.

pattern along the sperm flagellum. Contemporary studies have adopted Fourier analysis to ascertain primary parameters, especially under distinct viscosity scenarios [5,31]. A prevalent observation is the formation of sperm bundles, which might enhance the sperm cell's efficacy and is suspected to be cooperative behavior of the sperm cell. The intricate processes of this bundling have been explored extensively by Rothschild [25], emphasizing how sperm metrics can vary during different stages of the bundling process [25].

In the quest for analytical precision, computer-aided automation in bull sperm cell analysis was pioneered by Farrell et al. [4]. Leveraging advanced computational and imaging techniques, this method offers a detailed and accurate evaluation of sperm cell movement and morphology. This transition to computer-assisted systems ensures standardized and reproducible results, positioning them as invaluable assets in clinical settings for fertility evaluations and in research environments demanding detailed kinematic data.

3. Implementation

The objective is to analyze videos of beating bull sperm flagella with the aim of extracting insights regarding flagella motion, as depicted in the schematic diagram in Fig. 1. Typical output parameters of interest are curvature, amplitude, wavelength, period time, phase shift and initial orientation. From these videos capturing the beating patterns, data regarding the positions of the sperm head and the flagellum are

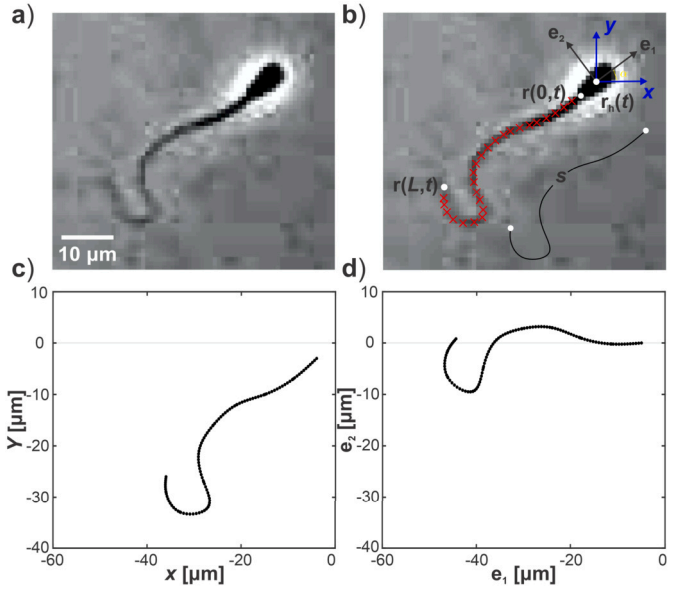

Fig. 2. Video of a sperm cell a) video frame at one time instant, b) head position $\mathbf{r_h}(t)$ (white dot) and flagellum shape $\mathbf{r}(s,t)$ (red crosses) are automatically detected in lab system $\mathbf{x} - \mathbf{y}$ using video analysis. c) Flagellum shape $\mathbf{r}(s,t)$ in lab system $\mathbf{x} - \mathbf{y}$. d) Flagellum shape $\mathbf{r_f}(s,t)$ in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$. This detection is done for all video frames for several beat cycles to obtain experimental input data.

extracted as input data using Matlab image analysis code (Fig. 2). To evaluate the methodology, validation data is generated using typical parameters and then tested to recover these parameters, thereby affirming the reliability of the applied techniques. As illustrated in Fig. 1, the input data is subjected to analysis through three distinct methods: Fourier analysis and two genetic GA. The resulting output parameters are compared with known parameters for the validation dataset. For experimental data, the parameters obtained through different methods are juxtaposed. Furthermore, the fitted shape of flagellar beating is compared to the input data, with a fit error serving as a metric to gauge the quality of fitting.

3.1. High-speed video microscopy of bovine spermatozoa

Cryopreserved bovine spermatozoa were received from Semex, Guelph, Canada, thawed for 2 minutes in a 37 degree water bath and then suspended in 2 mL high glucose Dulbecco's modified Eagle's medium with 4.5 g/L glucose, sodium pyruvate and sodium bicarbonate added. High viscosity medium was emulated by addition of 1% methylcellylose (Sigma Aldrich Nr. M0512) to the complete DMEM medium. The cell suspension was washed by centrifugation at 300 g for 5 minutes, removing the supernatant and resuspending the cells in clean medium. Videos were recorded at room temperature and atmospheric pressure immediately after washing with 100 frames per second for 500-1000 frames with a 10× or 20× objective in phase contrast mode with a Zeiss Axio Observer and Basler highspeed camera. 20 Microliters of cell suspension was placed on $0.18 \times 60 \times 24$ mm^3 microscope slides with coverslides attached by parafilm to form microchambers. The data of 22 sperm cells were collected from 13 videos from 8 different bull semen samples from different bulls. All sperm cells show normal swimming behavior as non-capacitating medium is used to avoid inducing capacitation.

3.2. Input data generation

The input data consist of two key components: the sperm head position versus time, denoted as $\mathbf{r_h}(t)$, and the flagellum's shape in the

laboratory reference frame, represented as $\mathbf{r}(s,t)$. These data sets are obtained through two means: they are either extracted from experimental video data (as depicted in Fig. 2) or calculated using equations, primarily for validation purposes.

3.2.1. Experimental data

Sperm cells consist of two primary components: an ellipsoidal head with a long axis of length $2a$ and a short axis of length $2b$, and a flagellum characterized by its length L and diameter D, as depicted in Fig. 2a. In experimental setups conducted between parallel glass plates, the flagellum predominantly deforms parallel to the surface of the glass. The analysis involves studying the flagellum beating of individual sperm cells S1 - S20, each swimming independently. Additionally, two sperm cells are part of a sperm bundle, referred to as B1 and B2. Consequently, video microscopy captures the primary beating motion of the flagellum. Automatic video analysis is employed to extract both the temporal changes in the head position, denoted as $\mathbf{r_h}(t)$, and the spatiotemporal changes in the flagellum, represented as $\mathbf{r}(s,t)$, within the laboratory coordinate system (lab system) denoted as $\mathbf{x}-\mathbf{y}$. Here, the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ of the $\mathbf{x}$- and $\mathbf{y}$-axes align with the orientation of the glass plates, while the $\hat{\mathbf{k}}$ vector is perpendicular to the glass plates. The path variable s, which ranges from 0 to L, traces the length of the flagellum, extending from the proximal end near the sperm cell head ($s=0$) to the distal end ($s=L$). The shape of the flagellum is available for up to 5 periods, with data captured at varying time points, ranging from 56 to 200 points, across different frames. Furthermore, there are 15 to 101 data points along the length of the flagellum for each analyzed sperm cell.

The flagellar shape is transformed from the laboratory system to the material system. The material system of the sperm cell is a coordinate system anchored at the head position of the sperm cell and is defined by two unit vectors: $\hat{\mathbf{e}}_\mathbf{1}$ aligns with the long axis of the sperm head, and $\hat{\mathbf{e}}_\mathbf{2}$ is perpendicular to it, as shown in Fig. 2. The determination of these unit vectors can be achieved through the following process:

$$\hat{\mathbf{e}}_\mathbf{1}(t)=\frac{\mathbf{r_h}(t)-\mathbf{r}(0,t)}{|\mathbf{r_h}(t)-\mathbf{r}(0,t)|},\ \text{and}\ \hat{\mathbf{e}}_\mathbf{2}(t)=\hat{\mathbf{k}}\times\hat{\mathbf{e}}_\mathbf{1}(t). \tag{1}$$

The flagellum deformation in material frame of reference is

$$\mathbf{r_f}(s,t)=((\mathbf{r}(s,t)-\mathbf{r_h}(t))\cdot\hat{\mathbf{e}}_\mathbf{1}(t),(\mathbf{r}(s,t)-\mathbf{r_h}(t))\cdot\hat{\mathbf{e}}_\mathbf{2}(t)). \tag{2}$$

3.2.2. Validation data

The flagellum beating is typically summarized using the following parameters: mean curvature K_0, bending amplitude A_0, wavelength λ, and period time T, as commonly employed in flagellum analysis [5]. In addition to these parameters, two more are considered: the initial phase shift ϕ_0 and the initial swimming direction α. These parameters are used to calculate the angle ϕ between the flagellum vector $\mathbf{r_f}$ and the negative direction of the long head axis $-\hat{\mathbf{e}}_\mathbf{1}$ through the following formula:

$$\phi(s,t)=K_0 s+A_0 s\sin\left(\frac{2\pi}{\lambda}s-\frac{2\pi}{T}t+\phi_0\right). \tag{3}$$

These additional parameters, such as the initial phase shift ϕ_0 and the initial swimming direction α, hold significance in characterizing sperm-sperm interactions, particularly in cases involving sperm bundling. The flagellum beating in the material frame of reference is described by the following formula:

$$\mathbf{r_f}(s,t)=-a\hat{\mathbf{e}}_\mathbf{1}-\int_0^s\left[\cos\phi(\tilde{s},t)\hat{\mathbf{e}}_\mathbf{1}+\sin\phi(\tilde{s},t)\hat{\mathbf{e}}_\mathbf{2}\right]\mathrm{d}s. \tag{4}$$

The resistive-force theory is applied to determine the head velocity $\mathbf{v_h}(t)$, which can be decomposed into its components along the material frame axes as $v_{\mathrm{h}1}(t)\hat{\mathbf{e}}_\mathbf{1}$ and $v_{\mathrm{h}2}(t)\hat{\mathbf{e}}_\mathbf{2}$. Additionally, the angular velocity $\Omega_h(t)$ is calculated using force and torque balance equations. These equations are instrumental in understanding the dynamic behavior of the sperm head and its motion, and we have

$$\zeta_1 v_{\mathrm{h}1}\hat{\mathbf{e}}_\mathbf{1}+\zeta_2 v_{\mathrm{h}2}\hat{\mathbf{e}}_\mathbf{1}+\int_0^s\left(\zeta_n\mathbf{v_n}+\zeta_p\mathbf{v_p}\right)\mathrm{d}s=0 \tag{5}$$

$$\zeta_{\mathrm{rot}}\Omega_{\mathrm{h}}(t)\hat{\mathbf{k}}+\int_0^s\mathbf{r_f}\times\left(\zeta_n\mathbf{v_n}+\zeta_p\mathbf{v_p}\right)\mathrm{d}s=0, \tag{6}$$

In accordance with [1], we define a set of friction coefficients that govern various aspects of sperm cell motion. Specifically, ζ_1 and ζ_2 represent the friction coefficients responsible for motion along the $\hat{\mathbf{e}}_\mathbf{1}$ and $\hat{\mathbf{e}}_\mathbf{2}$ directions, respectively, pertaining to the elliptical head. Additionally, ζ_{rot} accounts for the friction coefficient associated with rotational motion. Further, ζ_p is the friction coefficient governing translation of the flagellum parallel to its axis, while ζ_n influences translation normal to the flagellum axis. These coefficients play a pivotal role in characterizing the resistive forces experienced by the sperm cell during its motion, serving as key parameters in the resistive-force theory [1].

The velocities parallel and normal to the flagellum axis are determined as follows: First, the velocity vector $\mathbf{v_f}$ is obtained as the derivative of the flagellum position vector $\mathbf{r_f}$ with respect to time, i.e., $\mathbf{v_f}=\mathrm{d}\mathbf{r_f}/\mathrm{d}t$. Then, a tangent unit vector $\hat{\mathbf{t}}$ is computed as the derivative of $\mathbf{r_f}$ with respect to the path variable s, normalized by its magnitude, resulting in $\hat{\mathbf{t}}=\left(\mathrm{d}\mathbf{r_f}/\mathrm{d}s\right)/\left|\mathrm{d}\mathbf{r_f}/\mathrm{d}s\right|$. Using these vectors, the velocities parallel and normal to the flagellum axis are calculated as $\mathbf{v_p}=\left(\mathbf{v_f}\cdot\hat{\mathbf{t}}\right)\hat{\mathbf{t}}$ and $\mathbf{v_n}=\mathbf{v_f}-\mathbf{v_p}$, respectively. Subsequently, in the laboratory system, the initial ideal flagellum is rotated counterclockwise by an angle α. This rotation operation enables the description of flagellum motion in a reference frame that accounts for the initial swimming direction α

$$\mathbf{r}(s,0)=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix}\mathbf{r_f}(s,0), \tag{7}$$

The subsequent positions of the flagellum are determined from

$$\frac{\mathrm{d}\mathbf{r}(s,t)}{\mathrm{d}t}=\mathbf{v_h}(t)+\Omega_{\mathrm{h}}(t)\hat{\mathbf{k}}\times\mathbf{r_h}(s,t)+\mathbf{v_f}(s,t). \tag{8}$$

Validation data is generated using two sets of parameters, corresponding to low and high level of viscosity conditions, typical for bull sperm cells (cell medium and methyl cellulose enriched medium mimicking cervical mucus conditions), as detailed in Table 1 [5]. Ideal data is produced based on equations (3) through (8). To mimic the uncertainties associated with extracting input data from the video, scattered input data is created. At each position and time point, a random angle is added, following a normal distribution with a mean of zero and a standard deviation of 10°. This introduces variability into the input data, resembling real-world conditions where measurements may exhibit some level of imprecision, and we have

$$\phi_{\mathrm{s}}(s,t)=\phi(s,t)+\mathcal{N}(0,10°). \tag{9}$$

This process introduces alterations to the scattered input data vectors $\mathbf{r_f}$, $\mathbf{r}$, and $\mathbf{r_h}$ using equations (4) through (8) (left column in Fig. 3).

3.3. Sperm cell data evaluation methods

Sperm cell analysis methods extract the output parameters $K_0, A_0, \lambda, T, \phi_0, \alpha$ from the input data $\mathbf{r_h}$ and $\mathbf{r}$. The following explanation outlines the two methods: Fourier analysis-based decomposition is presented first, followed by the introduction of the new method GA.

3.3.1. Fourier method

The spatiotemporal deformation of the flagellum is extracted based on Equation (4) for test data and Equation (2) for experimental data. Subsequently, Equation (4) is inverted to obtain $\phi(s,t)=\arctan(\mathrm{d}r_{f2}/\mathrm{d}r_{f1})$ from the position vector $\mathbf{r_f}=r_{\mathrm{f}1}\hat{\mathbf{e}}_\mathbf{1}+r_{\mathrm{f}2}\hat{\mathbf{e}}_\mathbf{2}$. This

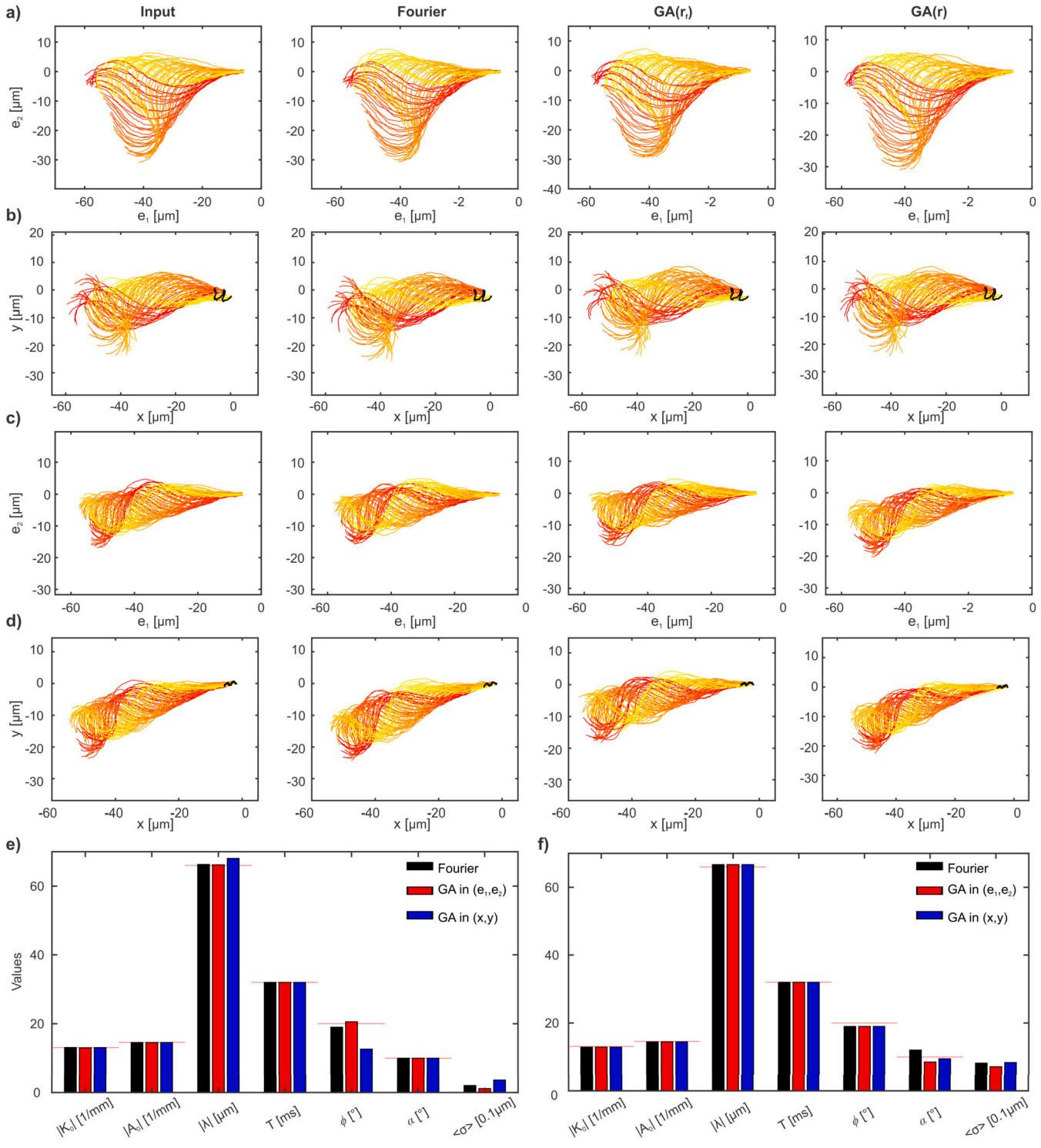

Fig. 3. Validation of data evaluation methods by usage of scattered test data. Left column: Scattered test data calculated by known parameters. Second column: Fourier analysis. Third column: results of GA($\mathbf{r_f}$) optimizing agreement of input data with simulated beating shape in material system. Right column: results of GA($\mathbf{r}$) optimizing agreement of input data with simulated beating shape in lab system. Beating shapes are shown for a) low viscosity in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$, b) low viscosity in lab system $\mathbf{x} - \mathbf{y}$, c) high viscosity in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$ and d) high viscosity in lab system $\mathbf{x} - \mathbf{y}$. Movement of proximal flagellum end $\mathbf{r}(0,t)$ (black). e) Comparison of achieved output parameters mean curvature K_0, beating amplitude A_0, wavelength λ, period time T, initial phase shift ϕ_0, initial swimming direction α and fit error $<\sigma>$ in comparison to input parameters (red dashed horizontal lines) for low viscosity, scattered input data. All methods Fourier analysis (black), GA($\mathbf{r_f}$) (red) and GA($\mathbf{r}$) (blue) give good agreement with input parameters. Fit error $<\sigma>$ is below 1 µm. Angles α (initial swimming direction) and ϕ_0 (initial phase shift) have the largest uncertainty.

inversion step is crucial for analyzing the flagellar deformation characteristics. The angle α represents the initial angle, defined at the proximal end where $s = 0$ and time $t = 0$, with respect to the **x**-axis. It is calculated as $\alpha = \arctan[(\mathrm{d}\mathbf{r}(0,0) \cdot \hat{\mathbf{j}})/(\mathrm{d}\mathbf{r}(0,0) \cdot \hat{\mathbf{i}})]$. The dominant peak in the Fast Fourier Transformation of $\mathbf{r_f}(0.5L,t)$ provides the inverse of the period time, denoted as $1/T$. The mean curvature K_0 is deter-

Table 1
Comparison of input validation parameters at low viscosity ($\eta = 0.7$ mPas) and at high viscosity ($\eta = 10$ mPas) with output parameters from input validation data using Fourier analysis, GA($\mathbf{r_f}$), and GA($\mathbf{r}$). The parameters are mean curvature K_0, beating amplitude A_0, wavelength λ, period time T, initial phase shift ϕ_0 and initial swimming direction α. The error relative to the input parameters is given in brackets. The fit error $\langle\sigma\rangle$ is $\langle\boldsymbol{\sigma_{12}}\rangle$ for methods **Fourier** and **GA($\mathbf{r_f}$)** and $\langle\boldsymbol{\sigma_{xy}}\rangle$ for **GA($\mathbf{r}$)**. Input validation parameters are almost perfectly recovered for all data evaluation methods. The highest deviations were found for scattered data at high viscosity in lab system. Angles α (initial swimming direction) and ϕ_0 (initial phase shift) have the largest uncertainty. Fit error $< \sigma >$ is below 1 μm. Scattered data increases fit error. Fit error is larger in lab system compared to material system, so that fit error $\langle\boldsymbol{\sigma_{xy}}\rangle$ for **GA($\mathbf{r}$)** is slightly larger than fit error of the other methods because of the included sperm cell motion.

	Method	K_0 [mm^{-1}]	A_0 [mm^{-1}]	λ [μm]	T [ms]	ϕ_0 [°]	α [°]	$\langle\sigma\rangle$ [μm]
Input	Parameter	13.1	14.6	66.0	32.0	20.0	10.0	-
Ideal	Fourier	13.10 (0.00%)	14.54 (0.43%)	66.31 (0.5%)	32.0 (0.0%)	19.0 (5.0%)	10.0 (0.0%)	0.20
low	GA($\mathbf{r_f}$)	12.99 (0.85%)	14.54 (0.43%)	66.21 (0.3%)	32.0 (0.0%)	20.6 (2.8%)	10.0 (0.0%)	0.12
viscosity	GA($\mathbf{r}$)	13.07 (0.21%)	14.54 (0.43%)	68.06 (3.1%)	32.0 (0.0%)	12.6 (36.8%)	10.0 (0.0%)	0.36
Scattered	Fourier	12.89 (1.58%)	14.47 (0.92%)	66.71 (1.1%)	32.0 (0.0%)	19.0 (5.0%)	12.0 (20.4%)	0.82
low	GA($\mathbf{r_f}$)	12.89 (1.58%)	14.47 (0.92%)	66.71 (1.1%)	32.0 (0.0%)	19.0 (5.0%)	8.5 (14.7%)	0.71
viscosity	GA($\mathbf{r}$)	12.89 (1.58%)	14.47 (0.92%)	66.71 (1.1%)	32.0 (0.0%)	19.0 (5.0%)	9.5 (5.2%)	0.84
Input	Parameter	6.6	12.0	39	54	20	10	-
Ideal	Fourier	6.60 (0.00%)	11.85 (1.22%)	39.09 (0.23%)	54.0 (0.00%)	19.0 (5.0%)	10.0 (0.0%)	0.19
high	GA($\mathbf{r_f}$)	6.56 (0.59%)	11.96 (0.37%)	39.13 (0.33%)	54.0 (0.00%)	21.0 (5.0%)	10.0 (0.0%)	0.11
viscosity	GA($\mathbf{r}$)	6.79 (2.86%)	11.70 (2.47%)	39.09 (0.23%)	54.0 (0.00%)	18.0 (9.8%)	8.9 (11.0%)	0.31
Scattered	Fourier	6.39 (3.14%)	12.18 (1.48%)	39.28 (0.73%)	54.0 (0.00%)	19.0 (5.0%)	12.0 (20.4%)	0.82
high	GA($\mathbf{r_f}$)	6.28 (4.91%)	11.82 (1.52%)	38.70 (0.76%)	54.0 (0.00%)	22.0 (10.0%)	4.7 (53.4%)	0.71
viscosity	GA($\mathbf{r}$)	9.29 (40.74%)	11.37 (5.24%)	37.45 (3.97%)	54.0 (0.00%)	27.9 (39.5%)	6.2 (38.5%)	1.00

mined as the slope of the time-averaged slope angle, denoted as $\langle\phi(s)\rangle$, which can be expressed as $\langle\phi(s)\rangle = \int_0^t \phi(s,t)\mathrm{d}\tilde{t}/t = K_0 s$. This mean curvature parameter is integral in characterizing the flagellar motion. The bending amplitude and wavelength are determined by fitting $\phi(s,t)$ at each position s with the function $2B(s)\cos(2\pi ft) + 2F(s)\sin(2\pi ft)$ to obtain the amplitudes $B(s)$ and $F(s)$. The bending amplitude A_0 is calculated as the slope of $\sqrt{B^2+F^2}$ with respect to s. The wavelength λ is found from the slope $2\pi/\lambda$ of $\arctan(F/B)$ as a function of s. These parameters, including $K_0, A_0, \lambda, T, \alpha$, are then employed to calculate the flagellum shape in both the material system $\mathbf{r_{fc}}(s,t)$ and the laboratory system $\mathbf{r_c}(s,t)$ using equations (3) through (8). In our case the fit error serves as a measure of agreement between the experimental data and the calculated data, reflecting the accuracy of the determined parameters in replicating the observed flagellar motion, meaning the fitness function of GA is maximized by minimizing fit error. The fit error in the material frame of reference is calculated as

$$\langle\sigma_{12}\rangle = \sqrt{\frac{\int_0^{t_f}\int_0^{L}\left[\mathbf{r_f}(s,t) - \mathbf{r_{fc}}(s,t)\right]^2 \mathrm{d}s\mathrm{d}t}{Lt_f}}, \tag{10}$$

and in the laboratory frame of reference

$$\langle\sigma_{xy}\rangle = \sqrt{\frac{\int_0^{t_f}\int_0^{L}\left[\mathbf{r}(s,t) - \mathbf{r_c}(s,t)\right]^2 \mathrm{d}s\mathrm{d}t}{Lt_f}}. \tag{11}$$

The phase shift ϕ_0 is adjusted iteratively until the fit error is minimized, signifying the alignment between the calculated and experimental data in the material frame of reference. This optimization process helps determine the optimal value of ϕ_0 that best represents the flagellar motion.

3.3.2. Genetic algorithm

The genetic algorithm is implemented using a population composed of N_P individuals, with each individual characterized by a gene consisting of six chromosomes (see Algorithm 1). This genetic representation allows for the exploration and optimization of the six parameters of interest within the population, and we have

$$G = (K_0, A_0, \lambda, T, \phi_0, \alpha). \tag{12}$$

In the GA implementation, the chromosomes represent the six unknown parameters. During initialization, each chromosome is randomly selected from an allowed range, as specified in Table 2. The bounds for these ranges are estimated based on parameters reported in the literature for different viscosities (see Table 1 and [5]). Notably, the maximum period time is extended to 600 ms because experimental data exhibited longer period times.

For each individual in the population, the fit error is determined based on its chromosome. This fit error quantifies the discrepancy between the input data's beating shape and the calculated flagellum shape. When conducting GA($\mathbf{r_f}$), the input flagellum shape is denoted as $\mathbf{r}_\mathrm{f}(s,t)$, and the calculated flagellum shape $\mathbf{r}_\mathrm{fc}(s,t)$ is obtained using the genetic algorithm's gene parameters through equations (12), (3), and (4). The fit error is computed using equation (10). GA($\mathbf{r_f}$) optimizes the agreement of the input data with the simulated beating shape in the material system represented by $\hat{\mathbf{e}}_\mathbf{1} - \hat{\mathbf{e}}_\mathbf{2}$.

Conversely, for method GA($\mathbf{r}$), the input data is $\mathbf{r}(s,t)$, and the calculated flagellum shape $\mathbf{r}_\mathrm{c}(s,t)$ is determined using equations (12), (3), and (8). The fit error is evaluated using equation (11). GA($\mathbf{r}$) aims to optimize the agreement between the input data and the simulated beating shape in the laboratory system denoted by $\mathbf{x} - \mathbf{y}$. The population is optimized over many generations N_G by search for lowest fit error. In each generation, mutation is happening to each chromosome with a probability of p_m. If the fit error of an individual decreases due to mutation, the old individual is replaced by the mutated individual. The crossover happens with probability p_c and exchanges randomly the chromosomes of two individuals. The old individual is replaced by the new individual in case of lower fit error. A selection process transfers the individual with lowest fit error to the next generation, repeatedly selects two individuals, transfers the individual with lower fit error to the next generation and fills a quarter of population with new random individuals.

4. Results

4.1. Method validation with test data

The methods have been thoroughly tested with validation data, as illustrated in Fig. 3 and summarized in Table 1. When ideal validation data is considered, there is an exceptional level of agreement between the input data and the calculated data, whether in the material or laboratory system and for both low and high viscosity conditions. The fit

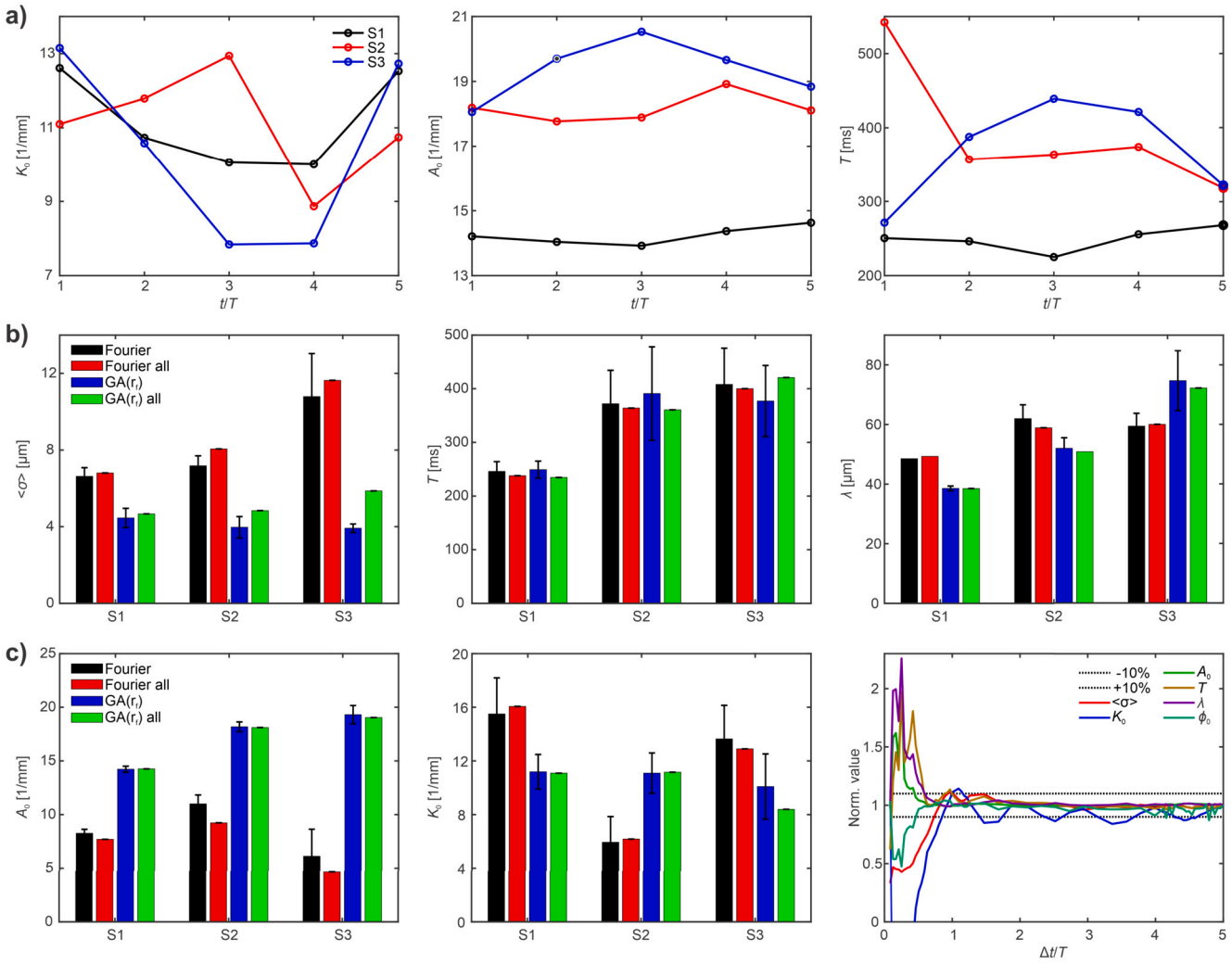

Fig. 4. Results for experimental input data from sperm cells S1-S3. a) Changes of mean curvature K_0, beating amplitude A_0 and period time T versus time for 5 period times for S1 (black), S2 (red), S3 (blue) using method **GA($\mathbf{r_f}$)**. b) Comparison of the parameters fit error $\langle\sigma\rangle$, period time T, wavelength λ, beating amplitude A_0, mean curvature K_0, achieved for Fourier method and GA($\mathbf{r_f}$), either averaged for individually analyzed separate 5 periods or for total 5 periods (all). Fit error is 4 – 12 µm and increases slightly if data from 5 periods is included instead of one period. **GA($\mathbf{r_f}$)** has smaller fit error than **Fourier** method. Period times are in the same range for all methods. Beating amplitude achieved by **GA($\mathbf{r_f}$)** is larger than by **Fourier** method. c) Normalized parameters for different time intervals by **GA($\mathbf{r_f}$)**. If data from less than one period is included in the analysis, the achieved parameters vary strongly. More than one period should be included to get variations less than 10 %. Therefore, typically about three periods are included in further analysis.

Table 2
Genetic algorithm parameters.

Parameter	Symbol	Value
Generations	N_G	1000
Population size	N_P	20
Mutation probability	p_m	0.2
Crossover probability	p_c	0.1
Length of flagellum	L [µm]	60 – 65
Diameter of flagellum	D [µm]	1
Head length	$2a$ [µm]	5
Head width	$2b$ [µm]	2.5
Curvature	K_0 [1/mm]	(−30) – 30
Amplitude	A_0 [1/mm]	1 – 30
Wavelength	λ [µm]	30 – 100
Period time	T [ms]	20 – 600
Phase constant	ϕ_0 [°]	(−180 – 180
Initial orientation	α [°]	(−180) – 180

error remains impressively low, with an average of $\langle\sigma\rangle < 0.36$ µm, representing the distance between the validation and calculated data. Notably, the fit error is minimized for GA($\mathbf{r_f}$) and highest for GA($\mathbf{r}$). The higher fit error observed in GA($\mathbf{r}$) is attributed to the consideration of sperm cell motion, which introduces additional sources of deviation from the ideal shape and motion. Furthermore, the key parameters K_0, A_0, λ, and T are accurately recovered, with errors below 3.1%. However, angles α and ϕ_0 exhibit slightly less precision, with deviations of up to 37%. The maximum deviation observed is 37% for the angle ϕ_0 at low viscosity and 11% for angle α at high viscosity.

When examining validation data with added scattering to simulate uncertainty in flagellum position extraction from videos, the fit error increases, reaching up to 0.84 µm at low viscosity and 1 µm at high viscosity. This is expected as disturbances from the ideal shape by scattering lead to less accurate fit with ideal shape. Despite this increased variability, the parameters K_0, A_0, λ, and T remain relatively accurate, with errors below 1.58% at low viscosity and up to 41% at high viscosity.

The shapes of the flagellum obtained from all methods (Fourier, GA($\mathbf{r_f}$), and GA($\mathbf{r}$)) closely resemble the test data and parameters. Fit error increases for scattered data and if sperm cell motion is taken into consideration.

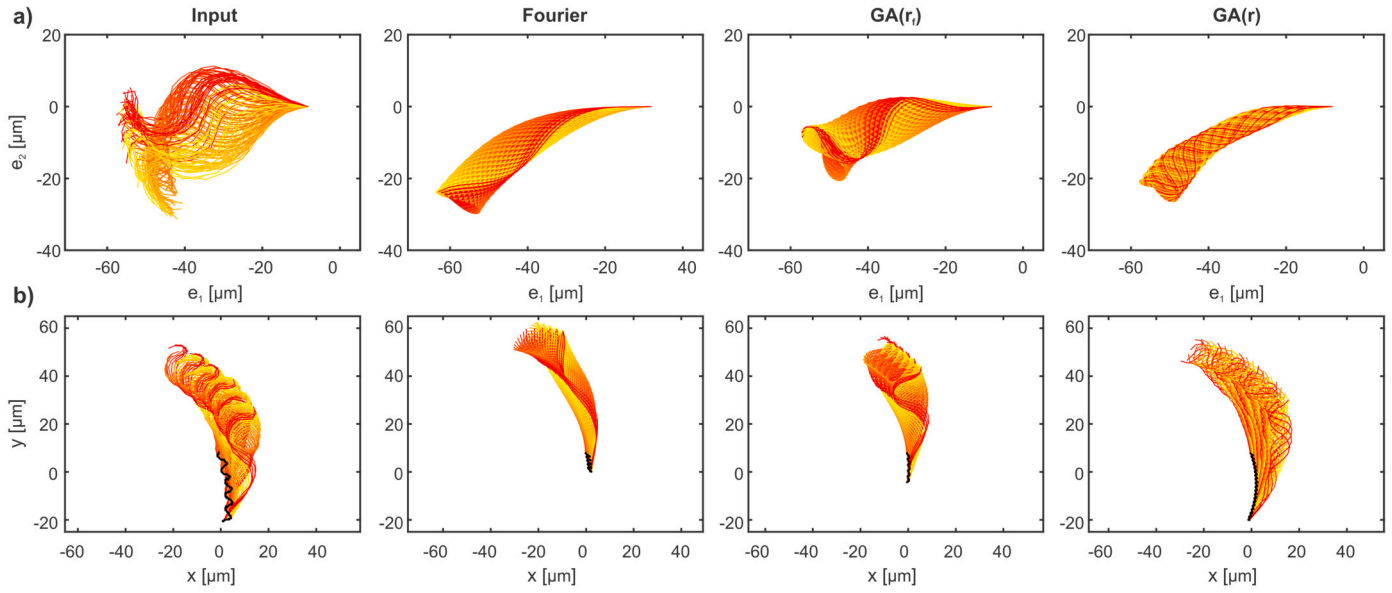

Fig. 5. Beating shapes for input data from sperm cell S1 and calculated from parameters a) in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$, b) in lab system $\mathbf{x} - \mathbf{y}$. Input data (left), **Fourier** (second column), **GA($\mathbf{r_f}$)** optimizes agreement of input data with simulated beating shape in material system (third column), **GA(r)** optimizes agreement of input data with simulated beating shape in lab system (right). Movement of proximal flagellum end $\mathbf{r}(0,t)$ (black). Beating shape and length, direction and curvature of proximal flagellum end movement are in good agreement for **GA(r)**.

4.2. Flagellum beating of individual sperm cells

We collected twenty datasets labeled S1 through S20, each corresponding to a distinct sperm cell exhibiting regular flagellum beating. These datasets encompassed flagellum coordinates, comprising approximately 15 to 100 points per time step, capturing the flagellum's motion over five periods, as depicted in Fig. 2. To extract parameters characterizing the sperm cell's motion, various methods were employed, as illustrated in Fig. 4. These methods included Fourier analysis for each period (referred to as "Fourier"). Subsequently, the results were averaged, and standard deviations were computed. Additionally, "Fourier all" entailed Fourier analysis using data from all five periods. We also utilized GA for each period, with the results of Fourier analysis serving as the starting point for 1000 generations. Similar to the Fourier approach, the results were averaged, and standard deviations were calculated. Finally, "GA all" employed GA with data from all five periods.

In Fig. 4a, we present plots of the mean curvature, bending amplitude, and period time determined by GA for the five periods. The mean curvature exhibited a range of $5-13$ mm^{-1}, the bending amplitude ranged from $14-19$ mm^{-1}, and the period time varied between $200-550$ ms. Notably, S1 displayed the smoothest trends, while S2 exhibited some sudden variations over time. The averaged parameters are visualized in Fig. 4b.

Comparing the three data sets, the error between measured and fitted shapes was higher using Fourier analysis than GA. GA reduced the error by an average of 32%, 45%, and 62% for sperm cells S1, S2, and S3, respectively, compared to Fourier analysis. Interestingly, the error using all data was slightly higher than using data from one period only. Regarding period T (beating period), S1 had the shortest period of about 238 ms, S2 had a medium period of 364 ms, and S3 had the longest period at 400 ms. All methods yielded approximately the same period of time, and the error bars were similar for both Fourier analysis and GA. S1 displayed the most regular beating and the smallest variation in period time. For S1 and S2, Fourier-based decomposition provided similar results for wavelengths, while GA yielded wavelengths that were 12% smaller for S1 and 14% smaller for S2. For S3, the wavelength obtained by GA was 22% larger than that obtained by Fourier analysis, with the smallest error bars observed for S1.

The bending amplitude A_0 was consistently larger when calculated using GA compared to Fourier analysis for all three sperm cells, with increases ranging from $72\%-85\%$ for S1, $65-97\%$ for S2, and a significant increase of $211\%-314\%$ for S3. The error bars were smallest for S1, followed by S2 and were largest for S3. Additionally, the error bars of GA were smaller than those of Fourier analysis. Mean curvature values fell within the range of $6-16$ mm^{-1}. GA produced a curvature 29% smaller for S1, 79% larger for S2, and 38% smaller for S3 compared to Fourier analysis. S1 and S2 displayed similar curvatures when using GA but exhibited significant differences when utilizing Fourier analysis. S2 had the smallest error bars, and in general, error bars for GA were equal to or smaller than those of Fourier analysis.

GA proved to be a superior method for achieving a better fit to experimental data compared to Fourier analysis, resulting in up to a 62% reduction in error. On average, the error ranged from $4-6$ µm instead of $7-12$ µm, potentially leading to a more accurate description of sperm cell behavior. Regarding the time interval for analysis, it was observed that there is no stable behavior if the time interval is shorter than one period time. The parameters stabilize to an error of about 10% when using at least one period time, but it is recommended to use at least two period times to obtain more stable results.

The beating shapes for the best parameters are plotted in Figs. 5–7. Each time the measured data is shown in the material system (sperm head is the origin of the coordinate system) and lab system. The motion of the sperm cell is visible in the lab system as the motion of the proximal end of the sperm flagellum is marked. Then, the bending shapes by Fourier analysis, GA($\mathbf{r_f}$) and GA($\mathbf{r}$) are shown. The best agreement with input data is typically for GA($\mathbf{r_f}$) in a material system.

The first sperm cell S1 beats asymmetrically (5a) and moves in negative y-direction (5b). In material system the beating shape of GA($\mathbf{r_f}$) is more similar to the input data than the beating shapes of Fourier analysis and GA($\mathbf{r}$). The width of the beating shape is the highest and there is widening at the end of the flagellum. This comes from the higher bending amplitude 15 mm^{-1} found by GA($\mathbf{r_f}$) compared to 7 mm^{-1} and 11 mm^{-1} (Table 3). The wavelength is found 40 µm and the period time 232 ms. The mean curvature is 11 mm^{-1} and smaller than with the other methods. In lab system, the beating shape of GA($\mathbf{r}$) has best agreement with the input data. The width is highest and the movement is similar to the input data. The initial orientation $-83-(-90)°$

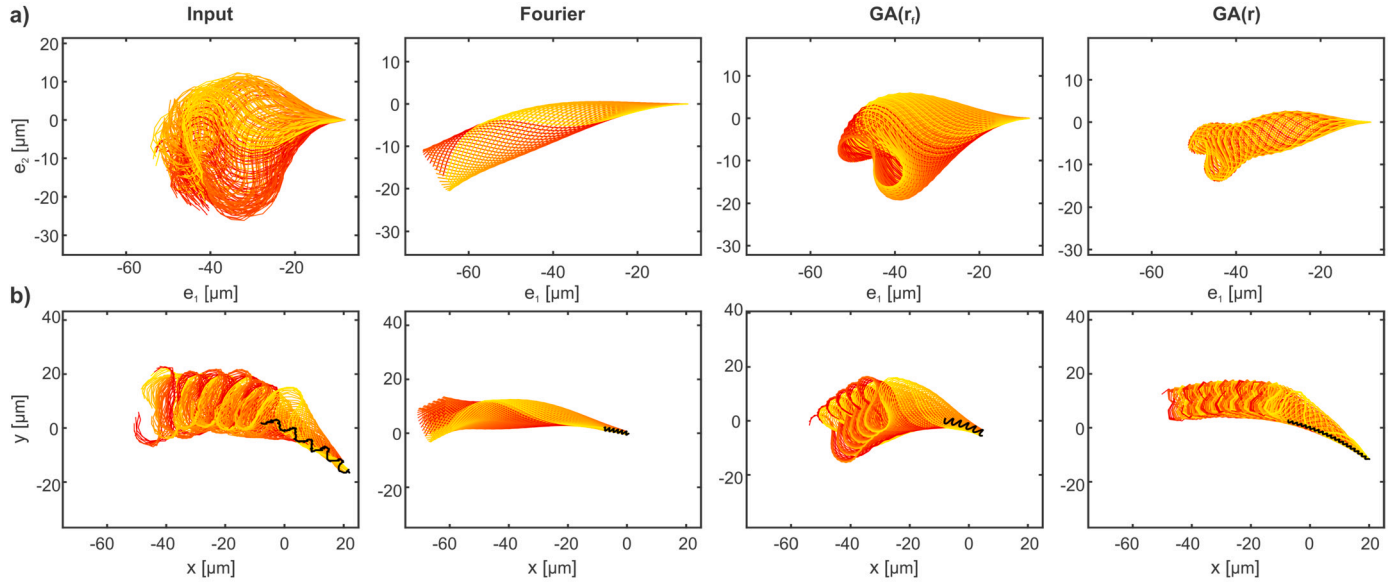

Fig. 6. Beating shapes for input data from sperm cell S2 and calculated from parameters a) in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$, b) in lab system $\mathbf{x} - \mathbf{y}$. Input data (left), **Fourier** (second column), **GA($\mathbf{r_f}$)** optimizes agreement of input data with simulated beating shape in material system (third column), **GA(r)** optimizes agreement of input data with simulated beating shape in lab system (right). Movement of proximal flagellum end $\mathbf{r}(0,t)$ (black). Beating shape and length, direction and curvature of proximal flagellum end movement are in good agreement for **GA(r)**.

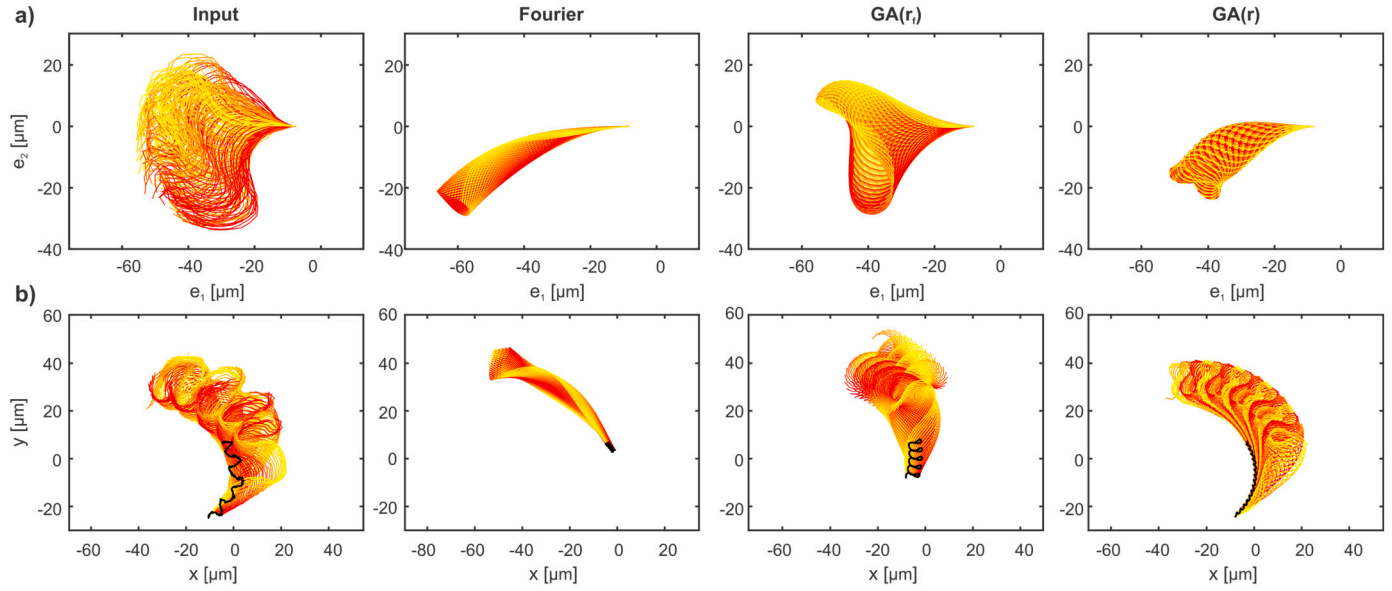

Fig. 7. Beating shapes for input data from sperm cell S3 and calculated from parameters a) in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$, b) in lab system $\mathbf{x} - \mathbf{y}$. Input data (left), **Fourier** (second column), **GA($\mathbf{r_f}$)** optimizes agreement of input data with simulated beating shape in material system (third column), **GA(r)** optimizes agreement of input data with simulated beating shape in lab system (right). Movement of proximal flagellum end $\mathbf{r}(0,t)$ (black). Beating shape and length, direction and curvature of proximal flagellum end movement are in good agreement for **GA(r)**.

agrees well with the movement in negative y-axis. Clockwise rotation during motion and beating shape having more values with negative y-values lead to positive mean curvature values $11 - 16\ \mathrm{mm}^{-1}$. The phase angle varies widely in the range $258 - 306°$. The fit error was successfully reduced from 7.4 µm for Fourier analysis to 4.3 – 4.6 µm for GA (reduction by 37% – 41%). The discrepancy between experimental data and the best fit obtained using the parameters and Eq. (3) is attributed to the high irregularities of sperm beating. While this may suggest that the basic model is inadequate, previous research [19] has demonstrated that parameters derived from the zeroth and first modes of Fourier decomposition encompass 95% of the power, rendering higher orders negligible. Fourier decomposition excels at detecting periodic behavior in scattered data, such as period time and wavelength. However, it does not optimize the complete 2D motion of the flagellum. Consequently, GA has the capacity to identify superior parameters by minimizing error.

Sperm cell S2 initially exhibits nearly parallel movement to the **x**-axis and gradually undergoes a clockwise rotation, ultimately aligning itself at an angle of approximately −45° relative to the **x**-axis at the end (see Fig. 6). Consequently, the angle of the initial orientation is relatively small, ranging from approximately −11° to −21°. In the material system, this sperm cell shows a more pronounced bending in the negative **y**-direction, resulting in a positive mean curvature of $8 - 11\ \mathrm{mm}^{-1}$ and clockwise movement. Fourier-based decomposition

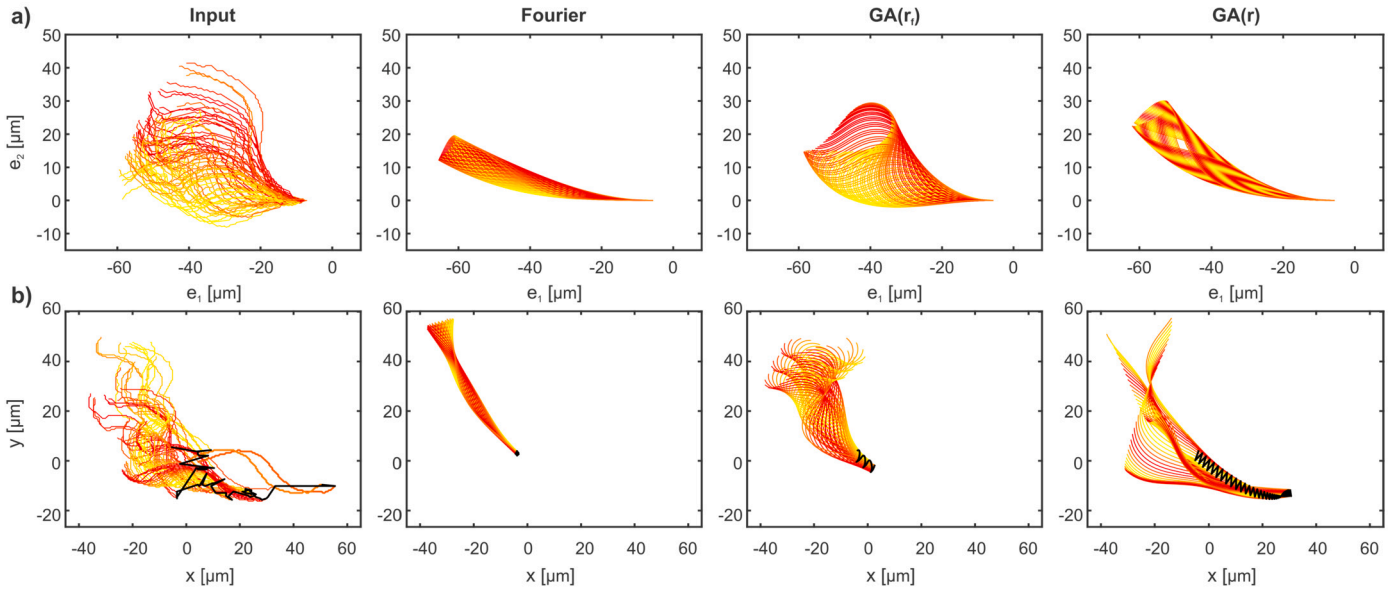

Fig. 8. Beating shapes for input data from sperm cell B1 and calculated from parameters a) in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$, b) in lab system $\mathbf{x} - \mathbf{y}$. Input data (left), **Fourier** (second column), **GA($\mathbf{r_f}$)** optimizes agreement of input data with simulated beating shape in material system (third column), **GA($\mathbf{r}$)** optimizes agreement of input data with simulated beating shape in lab system (right). Movement of proximal flagellum end $\mathbf{r}(0,t)$ (black).

identified a smaller bending amplitude of 6 mm^{-1}, while GA yielded a higher bending amplitude of 18 mm^{-1}, which better matched the bending shape observed in the input data. The wavelength was measured as approximately $51-52$ µm for both Fourier and GA($\mathbf{r_f}$), while GA($\mathbf{r}$) indicated a wavelength of 30 µm. The fit error ranged from 4.87 µm for GA($\mathbf{r_f}$) to 9.74 µm for Fourier analysis. This substantial reduction in the fit error of approximately 50% demonstrates the superior performance of GA in this context.

Sperm cell S3 initiates its movement at an angle of approximately $-70°$ relative to the x-axis and ends at around $-135°$ (as seen in Fig. 7). This pattern is also evident in the optimization results, as the initial angle α falls within the range of $-51°$ to $-63°$. In the material system, this sperm cell exhibits more pronounced bending in the negative y-direction, resulting in a positive mean curvature of $8-18$ mm^{-1} and clockwise movement. The movement, curvature, and bending shape are best captured using GA($\mathbf{r}$) in the laboratory system. The mean curvature is 18 mm^{-1} for GA($\mathbf{r}$), 8 mm^{-1} for GA($\mathbf{r_f}$), and 13.4 mm^{-1} for Fourier analysis. The bending amplitude is 17 mm^{-1} for GA($\mathbf{r}$), 19 mm^{-1} for GA($\mathbf{r_f}$), and 5 mm^{-1} for Fourier analysis. However, in the material frame of reference, the bending shape for GA($\mathbf{r}$) does not match the input data as well, while the width of the bending shape is a good fit for GA($\mathbf{r_f}$). Conversely, the movement does not match the input data as well in the laboratory system for GA($\mathbf{r_f}$). The fit error is reduced to 5.83 µm using GA($\mathbf{r_f}$) and 6.45 µm using GA($\mathbf{r}$) compared to 11.69 µm for Fourier analysis, representing a reduction of approximately $45\%-50\%$.

4.3. Flagellum beating in a bundle of cells

The data set B1 corresponds to 2.5 periods, as shown in Fig. 8, and parameters are summarized in Table 4. Using Fourier analysis, the results indicate negative curvature, very low bending amplitude, and an extremely large period time. The initial direction is approximately $-45°$, and the fit error averages 8.3 µm, significantly higher than the values obtained for validation data. When employing GA($\mathbf{r_f}$), higher mean curvature and bending amplitude magnitudes are obtained, while wavelength and period time remain similar to Fourier analysis. However, the phase shift differs by 20°. The average error is significantly reduced, dropping from 8.3 µm to 4.9 µm, which represents a substantial reduction of approximately 42%. GA($\mathbf{r}$) results in high curvature and low bending amplitude, with an unrealistically low period time compared to the other two methods. The phase shift also significantly deviates from the values found by Fourier analysis and GA($\mathbf{r_f}$). The initial orientation is close to 49°, similar to the 45° of the other methods. The fit error is reduced to 7.5 µm, representing a reduction of approximately 10%. In this case, GA($\mathbf{r_f}$) provides more reliable results compared to GA($\mathbf{r}$).

Sperm cell B2 exhibits a wide deformation and moves almost parallel to the $\mathbf{y}$-axis, as shown in Fig. 9, and parameters are summarized in Table 4. The mean curvature is approximately $17-18$ mm^{-1} for both Fourier analysis and GA($\mathbf{r_f}$), but it is smaller, around 9.4 mm^{-1}, for GA($\mathbf{r}$). Fourier analysis results in the smallest bending amplitude, while the other methods yield higher values, around $15-16$ mm^{-1}. The wavelength varies from 33 µm for GA($\mathbf{r}$) to 69 µm for GA($\mathbf{r_f}$) and up to 78 µm for Fourier analysis. The shape found by GA agrees better with the input data than the shapes obtained by Fourier analysis. The period times for Fourier analysis and GA($\mathbf{r_f}$ agree well, both around $560-600$ ms. The phase angle varies between $247°-300°$. The initial direction is approximately 90°, as found by Fourier analysis and GA in the material system. The fit error is reduced from the highest value of 9.8 µm with Fourier analysis to $6.5-6.9$ µm with GA, representing a reduction of approximately $29-34\%$.

4.4. Comparison

All parameters and fit errors are compared in Fig. 10 and Tables 3–4. It is evident that the fit error for validation data is significantly smaller than for experimental data. The smallest error is observed for the ideal data, while scattered data exhibits slightly higher errors but remains below 1 µm. In contrast, the fit error for real experiments exceeds 2.2 µm and can reach as high as 14 µm.

Comparing the methods, GA generally results in $10-70\%$ lower fit errors than Fourier analysis. For GA($\mathbf{r_f}$), the fit error is on average $50\pm 10\%$ of the fit error by Fourier decomposition. In approximately 68% of experiments, GA($\mathbf{r_f}$) yields smaller fit errors compared to the laboratory system. When considering period time, it is notably well recovered for validation data. Period time from Fourier analysis and GA($\mathbf{r_f}$) shows good agreement for all experimental data, with a maximum deviation of 10.2%. The lowest deviation is observed for S6 and S17. GA($\mathbf{r_f}$) is

Table 3
Output parameters for individual sperm cells (S1-S20) analyzed by **Fourier**, **GA($\mathbf{r_f}$)**, and **GA(r)**. Parameters: mean curvature K_0, beating amplitude A_0, wavelength λ, period time T, initial phase shift ϕ_0, initial swimming direction α, and fit error $< \sigma >$. Fit error $\langle \boldsymbol{\sigma} \rangle$ is $\langle \boldsymbol{\sigma_{12}} \rangle$ for **Fourier** and **GA($\mathbf{r_f}$)**, and $\langle \boldsymbol{\sigma_{xy}} \rangle$ for **GA(r)**. **Fourier** outputs serve as ground truth. For **GA($\mathbf{r_f}$)**, the largest change is an increase in amplitude from 4.4 μm to 15.4 μm. For **GA(r)**, beating amplitude increases from 4.4 μm to 12.8 μm, and period time decreases from 267 ms to 118 ms. Fit error is lower for **GA** than **Fourier**, and lower for **GA($\mathbf{r_f}$)** than **GA(r)**.

Input data	Method	K_0 [mm^{-1}]	A_0 [mm^{-1}]	λ [μm]	T [ms]	ϕ_0 [^{o}]	α [^{o}]	$\langle \sigma \rangle$ [μm]
	Fourier	15.29	6.85	51.20	248.0	269.0	-89.5	7.37
S1	GA($\mathbf{r_f}$)	10.61	14.67	40.12	232.4	257.5	-89.5	4.64
	GA(**r**)	15.92	11.22	30.33	68.1	305.9	-83.1	4.33
	Fourier	7.66	6.04	52.90	330.0	86.0	-11.3	9.74
S2	GA($\mathbf{r_f}$)	10.70	17.98	50.60	353.8	180.2	-11.3	4.87
	GA(**r**)	8.17	17.54	30.09	97.5	146.5	-21.3	5.85
	Fourier	13.42	5.03	61.13	408.0	10.0	-50.9	11.69
S3	GA($\mathbf{r_f}$)	8.21	19.03	72.34	422.7	13.1	-50.9	5.83
	GA(**r**)	18.14	16.55	39.77	133.5	115.3	-63.1	6.45
	Fourier	-9.12	9.34	43.35	243.3	95.0	122.6	5.63
S4	GA($\mathbf{r_f}$)	-6.19	15.06	40.06	224.3	82.5	122.6	3.67
	GA(**r**)	-3.90	14.52	30.11	40.4	20.7	126.7	4.28
	Fourier	-13.43	6.98	50.96	300.0	194.0	-81.0	6.18
S5	GA($\mathbf{r_f}$)	-8.13	18.39	38.29	194.7	131.1	-81.0	2.15
	GA(**r**)	-22.74	18.84	61.14	255.5	58.5	-62.2	4.43
	Fourier	-16.62	9.30	53.50	300.0	237.0	-46.5	6.45
S6	GA($\mathbf{r_f}$)	-10.08	17.48	44.51	302.7	270.5	-46.5	3.93
	GA(**r**)	-12.90	16.91	46.66	168.1	167.6	-40.0	4.03
	Fourier	-31.93	2.58	68.54	200.0	305.0	-124.6	9.36
S7	GA($\mathbf{r_f}$)	-29.94	14.57	99.79	218.4	324.5	-124.6	4.83
	GA(**r**)	-26.72	11.00	99.60	75.9	345.4	-160.0	7.21
	Fourier	-30.99	5.32	91.65	250.0	345.0	54.9	8.42
S8	GA($\mathbf{r_f}$)	-29.63	17.56	83.40	238.0	317.3	54.9	4.36
	GA(**r**)	-9.35	13.45	40.26	55.19	182.3	-0.8	7.13
	Fourier	-27.66	4.79	73.99	250.0	252.0	60.8	8.43
S9	GA($\mathbf{r_f}$)	-29.32	17.29	75.88	226.1	164.4	60.8	3.87
	GA(**r**)	-10.22	11.66	30.05	39.9	248.1	23.2	5.72
	Fourier	-16.07	2.63	92.81	250.0	196.0	-124.0	6.94
S10	GA($\mathbf{r_f}$)	-25.74	11.47	76.57	230.59	143.3	-124.0	2.70
	GA(**r**)	-22.05	13.05	70.26	217.16	89.4	-127.9	3.57
	Fourier	-14.91	1.27	81.36	200	190.0	-69.3	8.79
S11	GA($\mathbf{r_f}$)	-27.15	11.57	74.84	220.4	259.4	-69.3	2.97
	GA(**r**)	-18.39	13.44	55.41	212.4	273.2	-96.0	3.52
	Fourier	-17.15	3.55	81.53	250.0	166.0	-55.3	7.02
S12	GA($\mathbf{r_f}$)	-27.36	12.95	77.22	238.1	117.6	-55.3	2.69
	GA(**r**)	-21.68	12.20	53.19	229.5	155.3	-61.3	2.35
	Fourier	-17.15	3.52	81.40	250.0	166.0	-53.3	7.08
S13	GA($\mathbf{r_f}$)	-27.24	13.05	75.58	239.8	127.7	-53.3	2.75
	GA(**r**)	-22.15	12.34	51.87	234.8	178.8	-60.9	2.19
	Fourier	25.64	6.54	112.91	210.0	148.0	150.8	8.81
S14	GA($\mathbf{r_f}$)	29.05	20.81	86.40	222.3	196.6	150.8	4.30
	GA(**r**)	29.96	19.84	99.62	210.4	148.8	150.4	5.03
	Fourier	28.27	5.55	90.72	200.0	67.0	81.5	9.86
S15	GA($\mathbf{r_f}$)	26.92	20.35	93.79	208.8	91.5	81.5	3.53
	GA(**r**)	13.94	15.51	38.86	58.9	351.2	138.2	7.07
	Fourier	22.56	3.50	149.23	252.5	169.0	-57.6	7.98
S16	GA($\mathbf{r_f}$)	28.61	13.95	90.35	239.4	162.5	-57.6	4.49
	GA(**r**)	7.30	12.36	37.05	64.4	0.8	-9.5	6.65
	Fourier	-14.41	1.32	52.31	193.3	34.0	77.5	8.61
S17	GA($\mathbf{r_f}$)	-25.47	13.67	77.12	192.2	315.3	77.5	3.72
	GA(**r**)	-12.23	13.26	30.33	61.8	82.3	33.5	5.10
	Fourier	18.26	-1.71	100.09	333.3	122.0	-133.4	11.58
S18	GA($\mathbf{r_f}$)	29.85	11.73	98.81	296.6	219.4	-133.4	6.67
	GA(**r**)	29.99	2.11	80.48	27.9	266.7	-137.6	4.15
	Fourier	18.79	1.06	117.28	333.3	164.0	-142.8	14.10
S19	GA($\mathbf{r_f}$)	29.90	12.58	99.27	343.1	3.0	-142.8	8.00
	GA(**r**)	24.29	4.02	99.52	53.2	225.8	-119.8	4.49
	Fourier	20.88	1.27	118.79	333.3	197.0	-147.1	12.08
S20	GA($\mathbf{r_f}$)	29.88	13.97	96.48	303.5	338.7	-147.1	6.92
	GA(**r**)	10.72	6.05	52.84	64.4	334.9	-126.6	5.29
	Fourier	19 ± 7	4.4 ± 2.6	81 ± 28	267 ± 58	171 ± 87	-32 ± 92	8.8 ± 2.2
Average	GA($\mathbf{r_f}$)	22 ± 9	15.4 ± 3.0	75 ± 21	257 ± 60	186 ± 101	-32 ± 92	4.3 ± 1.5
	GA(**r**)	17 ± 8	12.8 ± 4.6	54 ± 24	118 ± 80	185 ± 107	-35 ± 92	4.9 ± 1.5

Table 4
Output parameters for sperm in sperm bundle (B1-B2) analyzed by **Fourier**, **GA($\mathbf{r_f}$)**, and **GA(r)**. Parameters include mean curvature K_0, beating amplitude A_0, wavelength λ, period time T, initial phase shift ϕ_0, initial swimming direction α, and fit error $<\sigma>$. The fit error $\langle\boldsymbol{\sigma}\rangle$ is $\langle\boldsymbol{\sigma_{12}}\rangle$ for **Fourier** and **GA($\mathbf{r_f}$)**, and $\langle\boldsymbol{\sigma_{xy}}\rangle$ for **GA(r)**. **Fourier** outputs serve as the ground truth. Average output values are at the table's end. For **GA($\mathbf{r_f}$)**, the largest change is an increase in beating amplitude from 3.7 μm to 14.9 μm. For **GA(r)**, beating amplitude increases from 3.7 μm to 10.2 μm, and period time decreases from 513 ms to 70 ms. Fit error is lower for **GA** than **Fourier**, and typically lower for **GA($\mathbf{r_f}$)** than **GA(r)**.

Input data	Method	K_0 [mm^{-1}]	A_0 [mm^{-1}]	λ [μm]	T [ms]	ϕ_0 [^{o}]	α [^{o}]	$\langle\sigma\rangle$ [μm]
B1	Fourier	-8.71	3.76	59.97	466.7	82.0	-45.0	8.33
	GA($\mathbf{r_f}$)	-16.94	14.18	59.89	454.5	61.2	-45.0	4.85
	GA(**r**)	-15.29	4.86	61.97	39.5	123.1	-49.0	7.51
B2	Fourier	-17.29	3.72	78.18	560.0	247.0	90.0	9.75
	GA($\mathbf{r_f}$)	-17.86	15.66	69.09	600.0	276.0	90.0	6.92
	GA(**r**)	-9.42	15.47	32.70	99.7	299.7	56.6	6.45
Average	Fourier	13 ± 6	3.7 ± 0.0	69 ± 13	513 ± 66	165 ± 117	23 ± 95	9.0 ± 1.0
	GA($\mathbf{r_f}$)	17 ± 1	14.9 ± 1.0	64 ± 7	527 ± 103	169 ± 152	23 ± 95	5.9 ± 1.5
	GA(**r**)	12 ± 4	10.2 ± 7.5	47 ± 21	70 ± 43	211 ± 125	4 ± 75	7.0 ± 0.8

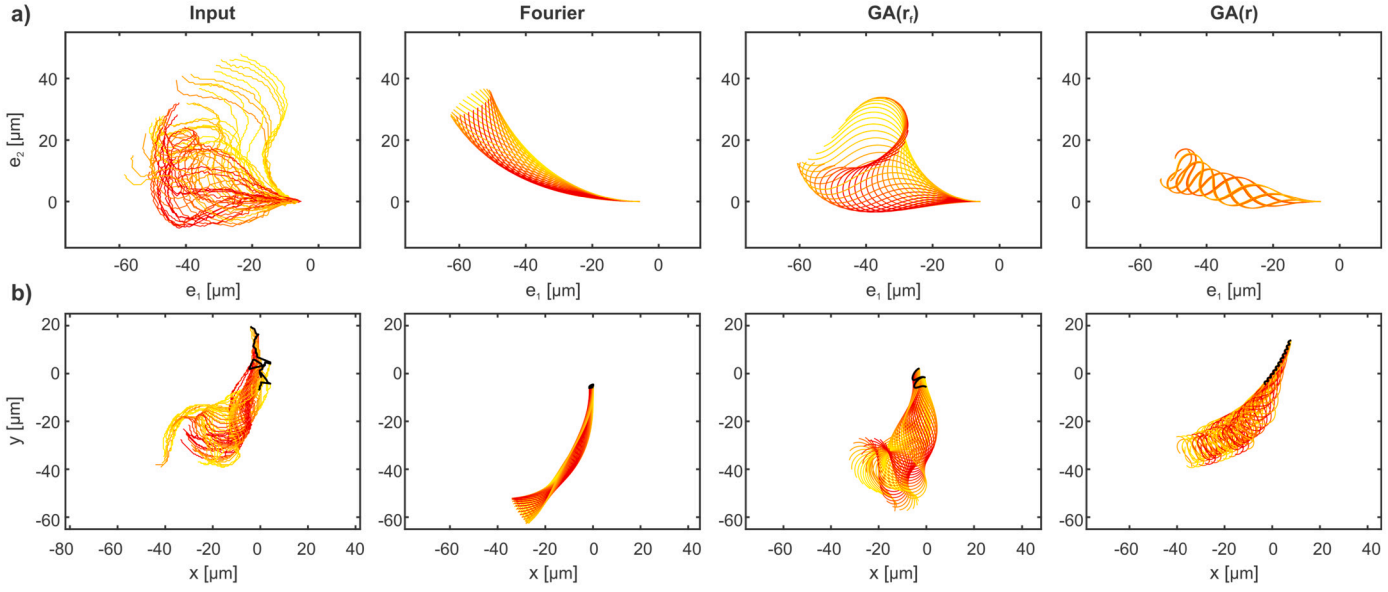

Fig. 9. Beating shapes for input data from sperm cell B2 and calculated from parameters a) in material system $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$, b) in lab system $\mathbf{x} - \mathbf{y}$. Input data (left), **Fourier** (second column), **GA($\mathbf{r_f}$)** optimizes agreement of input data with simulated beating shape in material system (third column), **GA(r)** optimizes agreement of input data with simulated beating shape in lab system (right). Movement of proximal flagellum end $\mathbf{r}(0,t)$ (black).

well-suited for determining the correct period time. In contrast, GA(**r**) underestimates the period time for 91% of experimental data, by an average factor of 2.3. This suggests that using experimental data from fewer periods may be necessary.

Regarding bending amplitude A_0, it is well-recovered for validation data (Fig. 10c). However, for experimental data, Fourier analysis tends to yield lower values, by a factor of 2-12, compared to GA($\mathbf{r_f}$). The higher bending amplitude achieved by GA($\mathbf{r_f}$) appears more reasonable, as observed in Figs. 5–7. This increased bending amplitude is a key factor contributing to the improved fitting error of GA compared to Fourier analysis. The curvature, as shown in Fig. 10d), displays no clear trend. In 36% of the cases, the curvature obtained by GA in the material system is smaller than that from Fourier analysis. Nevertheless, the sign of curvature remains consistent among the three methods. Wavelength agreement is generally good between Fourier analysis and GA($\mathbf{r_f}$), with an average deviation of 6 μm or 7.5% (Fig. 10e). However, GA(**r**) yields on average 34% smaller values for the wavelength than Fourier analysis. The standard deviation is consistent across all three methods for all parameters, except for an elevated standard deviation observed for amplitude in GA(**r**). This suggests that all three methods exhibit a similar range of uncertainty in their parameter estimations.

The initial orientation angle α is consistent between Fourier and GA($\mathbf{r_f}$) since the same method was used. Deviations in this angle are on average 21° ± 18°, with the highest deviation observed for S8 and S15. In contrast, the phase angle ϕ_0 exhibits similar trends among the three methods, but with larger deviations, ranging from 0.8 – 168° between the lowest and highest values of the three methods. This discrepancy is notably greater than the deviations observed for α, indicating that the orientation angle is determined with higher accuracy than the phase angle.

The validation data demonstrates excellent agreement for all three methods, affirming their effectiveness. It is important to note that curvatures can have positive or negative values, depending on whether the sperm cells exhibits clockwise or counter-clockwise bending. This is closely tied to the direction of the sperm cell's motion – whether it is swimming clockwise or counter-clockwise.

5. Conclusions

In the pursuit of enhanced analytical accuracy, GA presents a novel paradigm for analyzing sperm cell motility, offering a contrasting perspective to traditional Fourier analyses. Its primary advantage lies in its capability to juxtapose observed flagellar configurations directly against

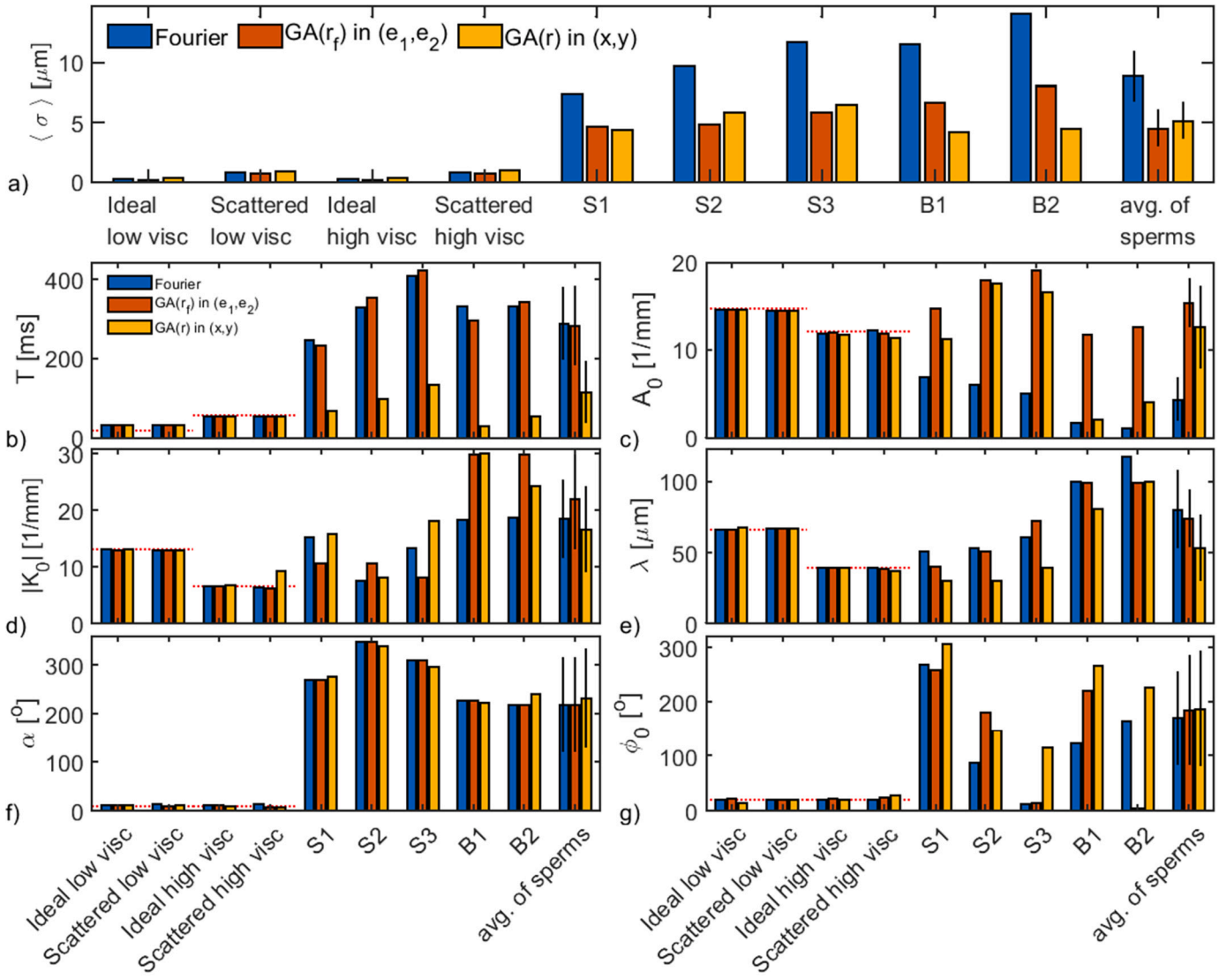

Fig. 10. Comparison of a) fit error, b) period time, c) bending amplitude, d) mean curvature, e) wavelength, f) initial orientation and g) phase constant for test data (ideal and scattered at low and high viscosity), five experimental data sets (S1, S2, S3, B1, B2) and average of 22 sperm cells analysed by **Fourier** (blue), genetic algorithm in material system **GA($\mathbf{r_f}$)** (red) and genetic algorithm in laboratory system **GA(r)** (yellow). Input parameters for validation data are indicated by red horizontal lines. Functionality of data evaluation methods was proven by low fit error for validation data. For experimental data, fit error is lowest for **GA($\mathbf{r_f}$)**, then **GA(r)** and highest for **Fourier**. Time periods found by **Fourier** and **GA($\mathbf{r_f}$)** agree well. **GA($\mathbf{r_f}$)** finds typically lower values for period time. Genetic algorithm methods find typically higher bending amplitude than **Fourier**. Other output parameters show similar range. Therefore, **GA($\mathbf{r_f}$)** is suggested as alternative or improvement for **Fourier**.

simulation outputs, leading to more accurate results of parameter estimations. This eliminates the need for intermediate data transformation steps that involve curve fitting and angle determination. Consequently, GAs can diminish fitting discrepancies by a remarkable 45.1% in contrast to Fourier analysis-based methodologies.

Additionally, GA allows for the determination of additional parameters, such as the phase constant and the initial orientation. These parameters can provide valuable insights, particularly in understanding the formation of sperm pairs. The detailed analysis of sperm motion and bundle formation in fluids with varying rheological properties requires further investigation. The method presented here sheds light on different flagellar behaviors in low and high viscosity environments. So far, no clear effect of increased viscosity on bundle formation has been observed. In future studies, we aim to explore the effects of fluid properties, such as elasticity versus viscosity of the medium, on sperm bundle velocity and flagellar waveform. Additionally, biochemical processes that alter sperm motion, such as hyperactivation and capacitation, are worth mentioning.

In our experiments, we intentionally used DMEM as a serum-free medium to prevent potential capacitation of sperm cells during measurements. DMEM was modified to include glucose and pyruvate as fuel sources to sustain sperm motility for several hours. In a previous study [17], we examined the impact of DMEM with and without serum in high and low viscosity on bundle formation and found these factors to have negligible effects on bundle occurrence. Detailed analyses of flagellar behavior, synchronization, and bundle velocity are still needed.

Fourier analysis, however, involves transferring sperm motion to the material system for flagellum deformation analysis, ignoring motion in the laboratory system. It is important to note that the inherent stochastic nature of GA requires more computational time and can produce variable outcomes across different runs. A useful strategy is to initialize one individual's gene in the population with results from Fourier analysis, jump-starting the optimization process and saving time. To reduce the randomness of results, increasing the number of generations or population size can lead to more consistent outcomes.

Currently, GA in the material system outperforms GA(**r**), especially in estimating the period time, which shows a high systematic error. To improve results, we might consider fixing the gene related to period time or using experimental data from fewer periods. Another potential enhancement could involve adjusting the fit error calculation to assign greater weight to points at the distal end of the sperm cell compared to the proximal end. Several parameters such as sperm cell dimensions and frictional coefficients were fixed in this study. These parameters could be included in the gene pool to improve the fitting. The observed fitting error appears significant relative to the length of the sperm cell, reflecting the presence of random fluctuations in experimental flagellar beating. While the current model captures 95% of flagellar power, alternative models for sperm cell beating may provide additional insights. Furthermore, factors such as sperm medium viscosity and capacitation status can now be examined using both methods. It remains crucial to determine which methods yield reliable results under specific conditions.

To enhance clinical applicability, using recording equipment with higher magnification than that used in conventional sperm cell analysis is essential. This would enable clearer visualization and detailed examination of sperm flagella, aiding in accurate analysis and diagnosis. Additionally, it is advisable to ensure that recorded videos contain only a limited number of non-rotating sperm cells. This simplifies the detection and tracking of individual sperm flagella, facilitating more efficient analysis and reducing potential errors in interpretation.

In general, GA and other artificial intelligence/deep learning methods offer a promising avenue for understanding sperm flagellum deformation, shedding light on sperm cell motion, and enhancing our comprehension of bundle formation. These techniques hold significant potential for advancing our knowledge in this field.

CRediT authorship contribution statement

Anke Klingner: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Alexander Kovalenko:** Writing – review & editing, Writing – original draft, Validation, Investigation. **Veronika Magdanz:** Writing – review & editing, Writing – original draft, Resources, Project administration, Funding acquisition, Formal analysis, Data curation. **Islam S.M. Khalil:** Writing – review & editing, Visualization, Validation, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgements

The collaboration project is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to the University of Twente to stimulate public-private partnerships (Grant number LSHI23002) and the Twente University RadBoudumc Opportunities (TURBO) program 2022. VM thanks NSERC Discovery grant and CFI for funding. The authors would like to thank Samantha Schwartz and Aaron Lewis for assistance with the videorecording.

References

[1] Bayly PV, Lewis BL, Ranz EC, Okamoto RJ, Pless RB, Dutcher SK. Propulsive Forces on the Flagellum during Locomotion of Chlamydomonas reinhardtii. Biophys J 2011;100:2716–25.

[2] Dai C, Zhang Z, Huang J, Wang X, Ru C, Pu H, et al. Automated non-invasive measurement of single sperm's motility and morphology. IEEE Trans Med Imaging 2018;37(10):2257–65.

[3] Das C, Mokashi C, Mande SS, Saini S. Dynamics and control of flagella assembly in Salmonella typhimurium. Front Cell Infect Microbiol 2018;8:36.

[4] Farrell PB, Presicce GA, Brockett CC, Foote RH. Quantification of bull sperm characteristics measured by computer-assisted sperm analysis (CASA) and the relationship to fertility. Biophys J 2011;49(4):871–9. https://doi.org/10.1016/S0093-691X(98)00036-3.

[5] Friedrich BM, Riedel-Kruse IH, Howard J, Jülicher F. High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory. J Exp Biol 2010;213:1226–34. https://doi.org/10.1242/jeb.039800.

[6] Gaffney EA, Ishimoto K, Walker BJ. Modelling motility: the mathematics of spermatozoa. Front Cell Dev Biol 2021;9:710825.

[7] Gallagher MT, Cupples G, Ooi EH, Kirkman-Brown JC, Smith DJ. Rapid sperm capture: high-throughput flagellar waveform analysis. Hum Reprod 2019;34(7):1173–85.

[8] Gray J. The movement of spermatozoa of the bull. J Exp Biol 1958;35:96–108.

[9] Hansen JN, Rassmann S, Jikeli JF, Wachten D. SpermQ–a simple analysis software to comprehensively study flagellar beating and sperm steering. Cells 2018;8(1):10.

[10] Hidayatullah P, Mengko TL, Munir R, Barlian A. Bull sperm tracking and machine learning-based motility classification. IEEE Access 2021;9:61159–70.

[11] Holwill MEJ. Physical aspects of flagellar movement. Physiol Rev 1966;46(4):696–785.

[12] Khalil ISM, Dijkslag HC, Abelmann L, Misra S. MagnetoSperm: A microrobot that navigates using weak magnetic fields. Appl Phys Lett 2014;104:22.

[13] Liu J, Leung C, Lu Zhe, Sun Yu. Quantitative analysis of locomotive behavior of human sperm head and tail. IEEE Trans Biomed Eng 2012;60(2):390–6.

[14] Magdanz V, Khalil ISM, Simmchen J, Furtado GP, Mohanty S, Gebauer J, et al. IRONSperm: Sperm-templated soft magnetic microrobots. Sci Adv 2020;6(28):eaba5855.

[15] Magdanz V, Schmidt OG. Spermbots: potential impact for drug delivery and assisted reproductive technologies. Expert Opin Drug Deliv 2014;11(8):1125–9.

[16] Montenegro-Johnson TD, Smith AA, Smith DJ, Loghin D, Blake JR. Modelling the fluid mechanics of cilia and flagella in reproduction and development. Eur Phys J E 2012;35:1–17.

[17] Morcillo i Soler P, Hidalgo C, Fekete Z, Zalanyi L, Khalil ISM, Yeste M, Magdanz V. Bundle formation of sperm: influence of environmental factors. Front Endocrinol 2022;13:957684. https://doi.org/10.3389/fendo.2022.957684.

[18] Raj B, Ahmedy I, Idris MYI, Noor RMD. A hybrid sperm swarm optimization and genetic algorithm for unimodal and multimodal optimization problems. IEEE Access 2022;10:109580–96.

[19] Riedel-Kruse IH, Hilfinger A, Howard J, Jülicher F. How molecular motors shape the flagellar beat. HFSP J 2007;1(3):192–208. https://doi.org/10.2976/1.2773861.

[20] Rikmenspoel R. Measurements of motility and energy metabolism of bull spermatozoa. Trans N Y Acad Sci 1964;26:1072–86.

[21] Rikmenspoel R. The inhibition by amytal of respiration and motiliy of bull spermatozoa. Exp Cell Res 1965;37:312–26.

[22] Rikmenspoel R. The tail movement of bull spermatozoa. Observations and model calculations. Biophys J 1965;5:365–92.

[23] Rikmenspoel R, van Herpen G. Cinematographic observations on the movements of bull sperm cells. Phys Med Biol 1960;5:167–81.

[24] Rikmenspoel R, Smits Utrecht HJ. Photoelectric and Cinematographic Measurements of the "Motility" of Bull Sperm Cells; 1957.

[25] Rothschild L. "The Cell and the OYgznism" ed. by J. A. Ramsay and V. B. Wigglesworth. Cambridge Univ. Press; 1961.

[26] Sakkas D, Ramalingam M, Garrido N, Barratt CLR. Sperm selection in natural conception: what can we learn from Mother Nature to improve assisted reproduction outcomes? Hum Reprod Updat 2015;21(6):711–26.

[27] Sohail A. Genetic algorithms in the fields of artificial intelligence and data sciences. Ann Data Sci 2023;10:1007–18. https://doi.org/10.1007/s40745-021-00354-9.

[28] Srairi F, Meguellati M, Saidi L, Djeffal F, meguellati F. Analytical modeling and optimization of new swimming microrobot design using genetic algorithm computations. In: 14th international conference on Sciences and Techniques of Automatic control & computer engineering - STA'2013; 2013. p. 265–8.

[29] Sun J, Garibaldi J, Hodgman C. Parameter estimation using metaheuristics in systems biology: a comprehensive review. ACM Trans Comput Biol Bioinform 2012;9(1):185–202.

[30] Xu H, Medina-Sánchez M, Magdanz V, Schwarz L, Hebenstreit F, Schmidt OG. Sperm-hybrid micromotor for targeted drug delivery. ACS Nano 2018;12(1):327–37.

[31] Zhang K, Klingner A, Le Gars Y, Misra S, Magdanz V, Khalil ISM. Locomotion of bovine spermatozoa during the transition from individual cells to bundles. Proc Natl Acad Sci 2023;120(3):e2211911120. https://doi.org/10.1073/pnas.2211911120.