V International Conference on Particle-based Methods – Fundamentals and Applications
PARTICLES 2017
P. Wriggers, M. Bischoff, E. Oñate, D.R.J. Owen, & T. Zohdi (Eds)

POTENTIAL FOR INTERACTIVE DESIGN SIMULATIONS IN DISCRETE ELEMENT MODELLING

DANIEL N. WILKE[1], NICOLIN GOVENDER[2,3,4], PATRICK PIZETTE[4] and RAJ K. RAJAMANI[5]

[1] University of Pretoria
Centre of Asset and Integrity Management, University of Pretoria, South Africa, 0086
wilkedn@gmail.com `http://www.up.ac.za/centre-for-asset-integrity-management`

[2] Research Center Pharmaceutical Engineering
Research Center Pharmaceutical Engineering (RCPE) GmbH Graz , Austria
govender.nicolin@gmail.com `www.rcpe.at`

[3] Council for Scientific and Industrial Research (CSIR)
Centre for High Performance Computing (CHPC), CSIR, Cape Town, South Africa
govender.nicolin@gmail.com `https://www.chpc.ac.za`

[4] IMT Lille Douai, Univ. Lille, EA 4515 - LGCgE - Laboratoire de Génie Civil et géoEnvironnement, départment Génie Civil & Environnemental, F-59000 Lille, France
patrick.pizette@imt-lille-douai.fr and http://www.lgcge.fr

[5] University of Utah
Metallurgical Engineering Department, University of Utah, USA
rajkrajamani@gmail.com `faculty.utah.edu`

Key words: DEM, Graphical Processing Unit, Interactive Design, Interactive Simulation, Tumbling Mill, Numerical Multi-Fidelity Models

Abstract. This study investigates the potential for combining lower fidelity models with high performance solution strategies such as efficient graphical processing unit (GPU) based discrete element modelling (DEM) to not only do simulations faster but differently. Specifically this study investigates interactive simulation and design for which the simulation environment BlazeDEM-GPU was developed that allows researchers and engineers to interact with simulations. The initial results prove to be promising and warranting extensive research to be conducted in future which may allow for the development of alternative paradigms.

In addition to the design cycle, the role that this interactive simulation and design will play in education is invaluable as an in-house corporate training tool for young engineers to actively train and develop understanding for specific industrial processes. This would also allow engineers to conduct just-in-time (JIT) simulation based assessment of

processes before commencing on actual site visits, allowing for shorter and more focussed site excursions.

1 INTRODUCTION

Discrete element modelling gained momentum from the mid 90s when small industrial scale simulations of a couple of thousand spherical particles in two dimensions were analysed on a more regular basis. A decade later discrete element modelling simulations allowed for moderate industrial scale simulations of a few hundred thousand spherical particles [1]. Over the last five years the landscape of large scale industrial simulations started to emerge with the utilization of the graphical processing unit (GPU) [2]. It is now common to simulate a couple of million polyhedral shaped particles and tens of millions of spherical particles [2] within a couple of days. In addition, moderate scale industrial simulations can now be modelled within hours on a workstation instead of weeks or months on a cluster merely a decade ago. Combining the advances in high performance GPU based discrete element modelling with sound computational complexity reduction techniques such as discrete element simulations with varying fidelity of the physics or the numerics [13] and reduced order modelling [3] allows for the possibility of design optimization or design modification for industrial problems that involve granular flow. In design optimization the distinction between the need for *accurate analyses* and *accurate enough analyses* are seldom made although they can have significant impact on solution times when properly applied. Initially in the design optimization process the analyses only need to be *accurate enough* to capture the correct trends regarding an analyses. As a design converges the need for *accurate analyses* sometimes increases and then not even always. By clearly making a distinction between these two types of analyses has led to significant improvements of multi-fidelity design optimization approaches over the last decade [6, 7, 8, 9]. Here, multi-fidelity is used in the sense to imply fidelity of the physics or numerical computational fidelity of a simulation.

More importantly this allows for a new paradigm in design optimization and design modification that is distinct from the conventional design cycle. The design optimization cycle remains characterised by either the *analyse-wait-modify-analyse cycle* or more recently with cloud computing the *batch analyse-wait-modify-batch analyse cycle* utilizing mostly high fidelity numerical simulations as most design engineers or researchers do not distinguish between accurate solutions and solutions that give accurate trends during the design process. Combining a proper understanding between *accurate analyses* and *accurate enough analyses* that captures the consistent trends with a computationally efficient solution strategy for the discrete element method (DEM) allows for quick turn around times on feedback to the user. By incorporating this additional computational savings naturally allows us to extend the domain of application towards interactive design simulations in which low fidelity numerical models can be computed efficiently to give consistent and responsive feedback to the user regarding the trends of various parameters during design modification. Consequently combining lower fidelity models with high performance GPU based DEM modelling is enabling a new and alternative paradigm denoted interactive simulation and design (ISD). The benefit of ISD is that engineers can explore the design domain independently to gain understanding of a process or even explore the design domain guided by a numerical optimization strategy which ultimately leads to better understanding of the optimal solution as opposed to the often black-box approach to design optimization in which the engineer is often detached in terms of understanding from the optimal solution.

BlazeDEM-GPU [2] is a GPU based DEM simulation environment designed with ISD in mind. BlazeDEM-GPU allows for changes to be made during a simulation in real time utilizing multiple GPUs to conduct a single simulation. This capability allows engineers to interactively engage with a simulation to study the effect of various model parameters. For example geometrical changes of the environment with which particles interact that includes the effect of lifter heights in a ball design as an example. Changes in the boundary conditions in steady state processes that include inter-particle cohesion or changes in flow rate of a continuous bulk material handling process. This would allow engineers to interact with the simulation to both quantitatively and qualitatively engage with a design problem or granular material process, allowing for formalized and intuitive understanding of the factors that influence granular flow to ultimately drive towards an enhanced understanding of optimal design solutions.

In addition to ISD, the role that this paradigm will play in education is invaluable as an in-house corporate training tool for young engineers to actively train and develop understanding for specific industrial processes. This also allows engineers to conduct just-in-time (JIT) simulation based assessment of processes before commencing on actual site visits, allowing for shorter and more focussed site excursions. Ultimately, this would allow experienced engineers to explore new processes and solidify the understanding behind experiential experience that is becoming ever more important in our dynamic modern environment.

2 EXPLANATORY EXAMPLE FOR NUMERICAL FIDELITY

As an explanatory example to explore the concepts behind numerical fidelity of a model, consider the coupled first order non-linear system of differential equations (Lokta-Volterra equations [10] given by

$$\frac{dx(t)}{dt} = \alpha x(t) - \beta x(t)y(t) \tag{1}$$

$$\frac{dy(t)}{dt} = \delta x(t)y(t) - \gamma y(t), \tag{2}$$

with two initial conditions on $x(0) = x_0$ and $y(0) = y_0$ that completes the formulation. For illustrative purposes consider the one dimensional least squares problem for which we choose $\beta = \frac{1}{3}$, $\delta = \gamma = 1$ and $x_0 = y_0 = 0.9$. The aim is to find α that best matches $x(t)$ and $y(t)$ after $t = 8$ seconds. We compute the desired response $\mathbf{d}^*$ for $\alpha^* = 0.8$ using the forward Euler integration scheme using 1000 equal time steps between 0 and 8 seconds. Hence, the solution $\mathbf{d}^*$ is known which allows us to quantify our predicted solutions using a numerically lower fidelity model. Given some estimated response $\mathbf{d}_m(\alpha)$ using a lower fidelity model that depends on the following model parameter α. The difference between the known response and the model response is then given by

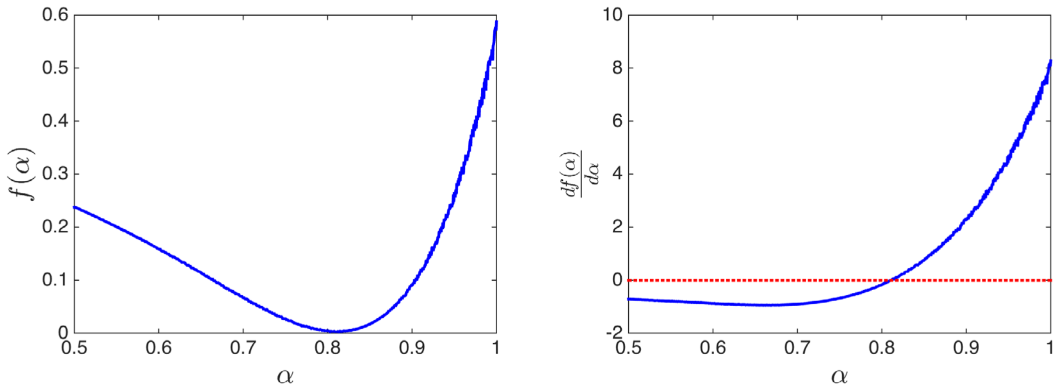
$$\mathbf{E}(\alpha) = (\mathbf{d}_m(\alpha) - \mathbf{d}^*), \tag{3}$$

which can be reduced to a scalar $f(\alpha)$ by the following ordinary least square projection

$$f(\alpha) = \mathbf{E}^T(\alpha)\mathbf{E}(\alpha). \tag{4}$$

Consider a lower numerical fidelity model using a random number of time steps between 290 and 350 for each computed α. By choosing a random number of integration time steps for each α allows us to quantify the variation in response between designs. The least squares error is depicted in Figure 1(a), while the derivative of the least squares error w.r.t. the variable α is depicted in Figure 1(b). It is evident that using a third of the computing power the estimated solution was resolved around 0.81 as opposed to $\alpha^* = 0.8$ for the known solution. In addition, the variation in least squares error when varying the number of integration time steps is negligible allowing for a smooth response in both the function value and derivative. Here, the function value is indicative of what is required when an accurate response is of importance for a design variable and of importance towards the end of a design optimization process. In turn, the derivative is indicative of the trend of a design variable and merely needs to point us in more or less the right direction which for the 1D case results in one of two options, left or right. A negative derivative is indicative of move right for improvement whereas a positive derivative is indicative of moving left for an improvement. For this example the two are equivalent and consistent pointing to the same solution.

However, by lowering the numerical fidelity of the computation to less than 10% of the known solution a clear distinction between the function value and derivatives becomes evident. Considering an even lower numerical fidelity model using a random number of time steps between 70 and 90 for each computed α. The least squares error is depicted in Figure 2(a), while the derivative of the least squares error w.r.t. the variable α is depicted in Figure 2(b). It is evident that using less than a tenth of the computing power the estimated solution from the function value is difficult to resolve. The variation between responses is significant as the number integration time steps between different values of α makes comparing the quality of solutions for different values of *salpha*. Hence, achieving an *accurate solution* is not achievable for this low fidelity computed solution. However, when we consider the trends of the solution as indicated by the derivative a unique solution around $\alpha = 0.83$ is evident as depicted in Figure 2(b). The estimated



(a) (b)

Figure 1: Solution of (a) a cost function and (b) the derivative of the cost function w.r.t. the design variable α. For each α a differential equation was required to be solved. Every α was solved using a random number of integration time steps between 290 and 350.

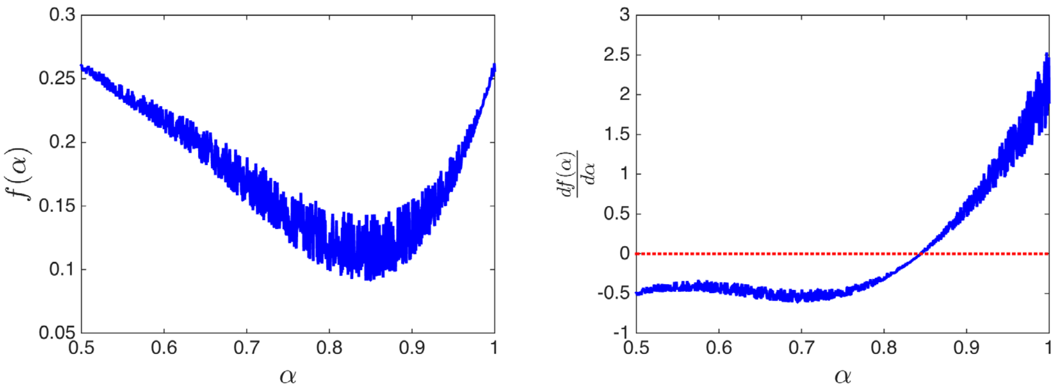
solution here is remarkably close to the actual solution of $\alpha^* = 0.8$ given we expended less than 10% of the computing power. The implication of this observation for solving a design optimization problem in general is that by initially focussing more on the trends of a simulation and only the values towards convergence if required can have a significant computational saving.

To reiterate the discussion in Section 1, design optimization remains characterised by either the *analyse-wait-modify-analyse cycle* or more recently the *batch analyse-wait-modify-batch analyse cycle* utilizing mostly high fidelity numerical simulations as most design engineers or researchers do not distinguish between accurate solutions and accurate trends in the design process. Combining lower fidelity models with high performance solution strategies such as efficient GPU based DEM enables us to not only do simulations faster, but differently. This allows for the development of new and alternative paradigms of which ISD is an example.

3 OVERVIEW OF COMPUTING ARCHITECTURES

The two most common computing platforms that are readily available are the central processing unit (CPU) and the graphical processing unit (GPU), while field programmable gate arrays (FPGAs) allow for increased computing power they are costly. In modern computing the following three criteria dominate the selection computing architectures:

1. time to solution,
2. energy required to compute a solution,
3. capital cost for computing platform.



(a) (b)

Figure 2: Solution of (a) a cost function and (b) the derivative of the cost function w.r.t. the design variable α. For each α a differential equation was required to be solved. Every α was solved using a random number of integration time steps between 70 and 90.

The energy requirements to compute a solution are becoming increasingly more important to consider as the number of problems being solved daily is growing substantially with the potential for significant power savings. However, for a given architecture the time to solution and power requirements are two competing objectives as exemplified by the power characteristics of CMOS integrated circuits (ICs). Both CPUs and GPUs follow the CMOS IC power characteristics that are dominated by the two main contributors namely, static and dynamic power draw. The static power consumption is the power required when the transistors are not switching which is given by

$$P_S = VI_S, \tag{5}$$

with V the supply voltage to the circuit and I_S the static current flowing though the circuit. The dynamic power which consists of capacitive and transient power consumption and follows

$$P_D = aC \times V^2 \times f, \tag{6}$$

with a representing the fraction of transistors switching, C indicating the switched capacitance, V the supply voltage and f the clock frequency [11]. The higher the clock frequency the lower the latency but the more power the processing consumes.

However, the time to solution and energy required to compute a solution is not directly related to the clock frequency when we consider different computing architectures as the potential for parallelization of the problem at hand is an important point to consider. The CPU and GPU are two computing architectures designed for two very different problems.

The central processing unit (CPU) is a contemporary general-purpose processor that excels at computing multiple instructions to be executed sequentially on new or the same data. Parallelization of CPUs have extended to only a few multi-core processing units

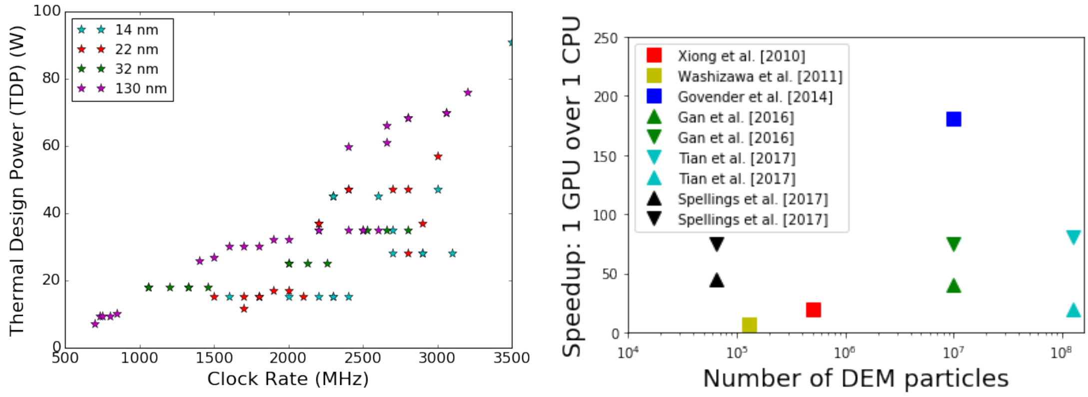


Figure 3: (a) Clock rate versus thermal design power for various transistor sizes of CPUs and (b) the speedup factor for 1 GPU over 1 CPU to compute the various number of particles.

(typically less than ≤ 32) with data being fed using limited memory bandwidth as a few cores can only compute so fast. These cores run at the highest clock cycles serving applications that require low latency at the cost of higher power demands. The relationship between clock rate as function of thermal design power (TDP) outlined by (6) is indicated in Figure 3(a) for various transistor sizes.

In turn, the graphical processing unit offers a vast number of simple, data-parallel and deeply multi-threaded cores that are fed data using high memory bandwidth. GPUs excel at problems that require the same operation to be performed on different data known as single instruction multiple data (SIMD) parallelization. The cores run at about one quarter ot third of the clock rate than the high-end CPUs reducing the power demand at the cost of sacrificing latency for throughput. The programming support for GPUs have improved significantly with CUDA supporting development for NVIDIA based GPUs while OpenCL renders support for AMDs manufactured GPUs. In addition, the GPU has significant speedup when conducting single precision floating point operations as opposed to double precision operations. Although, single precision DEM has only been explored by a limited number of researchers, the potential for additional computational performance and reduced memory requirements renders this an important consideration for future research. The reduced memory requirements allows for more particles to be solved using the same amount of memory which is important for memory limited computing platforms such as the GPU. In addition, the GPU is designed with a high bandwidth memory interface to allow large amounts of data to be moved in memory on the GPU for efficient large scale parallel processing.

DEM is ideally suited for GPU architectures as it is a throughput constrained computing problem, as a consequence it has resulted in significant speed-up over CPU based DEM simulations as depicted in Figure 3(b). The single precision performance by Govender et al. [12] indicates a significant improvement in computational performance of a factor of 180 when considering the GPU above the CPU, the general GPU performance improvement is around a factor of 50 to 100. Figure 3(b) clearly demonstrates the benefit of utilizing the GPU to solve DEM problems, since DEM problems with large numbers of particles are ideally solved by architectures designed for high throughput problems.

4 PRELIMINARY INVESTIGATION TUMBLING MILL DESIGN LANDSCAPE

This study investigates the potential for adaptive particle refinement to speed-up design based discrete element simulations by conducting a pilot investigation with the set of preliminary results reported in [13]. This study investigates the effect of the lifter geometry on the distribution of the estimated power draw of the mill with the estimated normal and shear energies to be considered in a future study. The two variables for the lifter geometry are the width and height as depicted in Figure 4(b). The height is varied 1.2cm and 4.8cm, while the widths are varied between 2cm and 10cm.

For the purposes of our investigation we conduct three sets of analyses using three particle sizes. We consider the finest particle size to be the solution landscape. We therefore compare the response surface landscapes not only w.r.t. to the predicted values but also the distance each response surface's optimal design is from the known solution. Three particles sizes selected are $r_1 = 1$ in, $r_2 = 0.794$ in and $r_3 = 0.63$ in. As a first approach the selection of the sizes are based on the following criteria:

1. The radii being related by $0.5^{\frac{1}{3}} \approx 0.794$ should fill the same representative volume when number of particles are doubled.

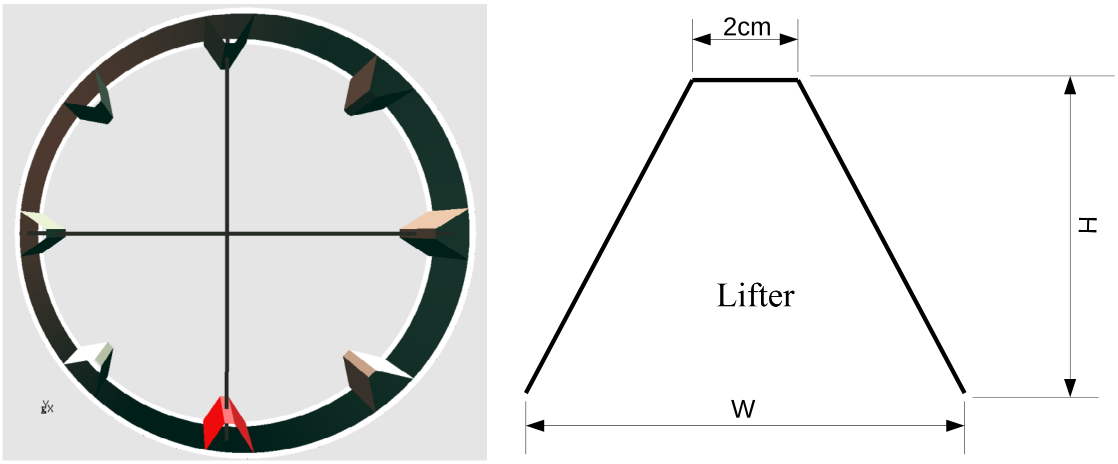
2. The potential energy of the system stays the same when the effective masses m and heights h stay the same mgh, with g the gravitational acceleration.

3. The kinetic energy, $\frac{1}{2}mv^2$, stays the same when the mass stays the same and the velocity stays the same. The tumbling mill is modelled with prescribed rotations that enforces the same velocities as input the problem.

Selecting the particles following the relationship $r_2 = 0.794r_1$ ensure the masses to be the same, particle volume to be the same, as well as the potential energy as the particles are loaded to the same height when the particle number is doubled and particle volume halved for respectively 1000, 2000 and 4000 particles. The volumes being effectively the same for the three particle sizes are evident in Figures 5(a)-(c). The impulses to the problem will vary significantly due to the lumped nature of the particles.

We consider BlazeDEM-GPU [12], where the equations of motion are integrated using an explicit forward Euler scheme using a time step $\Delta t = 10^{-4}$. The particle-particle coefficient of restitution is 0.4 and the stiffness chosen to resolve the contact within at least 10 time steps. The static and dynamic coefficient of friction is chosen as 0.45 and the rolling resistance is selected as 0.001. The particle-cylinder and particle-lifter properties coefficient of restitution is 0.4 and the stiffness chosen to resolve the contact within at least 10 time steps. The static and dynamic coefficient of friction between the particle-cylinder and particle-lifter is chosen as 0.5 and the rolling resistance is selected as 0.001. The particle density is assumed to be of Aluminum 2.7 cm/g^3.

After the 20 discrete element runs per particle diameter have been completed, as depicted in Figures 6(a)-(c), the radial basis function (RBF) response surfaces for each particle size was constructed as depicted in Figures 7(a)-(c). To construct the RBF surface the optimal ϵ was first determined by minimizing the leave-out-one-cross-validation error (LOOCVE) using a brute force strategy [13]. Once the optimal ϵ^* has been determined, it is used to construct the radial basis function (RBF) response surface.

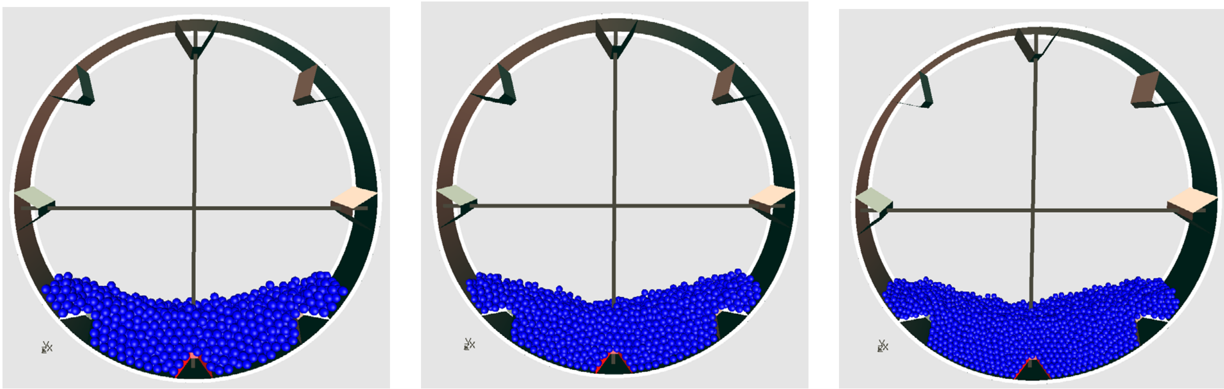


(a) (b)

Figure 4: (a) Modelled ball mill, with 90cm diameter, depth of 15cm and eight lifters, filled with spherical charge particles. (b) The geometry of the lifter parametrized using two variables [13].

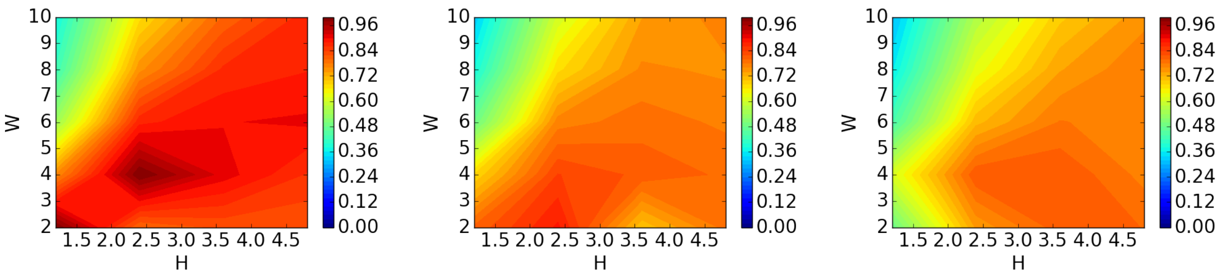
The RBF predicted normalized power draw is depicted in Figures 7(a)-(c) by dividing by the configuration with the highest power demand. It is evident when considering the domains with the highest reported power demand all three simulations are within the same part of the design domain. This demonstrates promising results for utilizing such responses as objective function in design optimization problem, as the actual values are not so important as the design vector that coincides with the maximum power requirements or minimum.

In turn, when considering the actual values as exemplified by the contours that only show values limited between 0.8 and 0.9 as depicted in Figure 8(a)-(c) it is clear that the 1000 particles simulation seems to be inadequate, whilst the 2000 particle simulation demonstrates promising potential. This is an important consideration when considering constraint functions as constraints require accuracy as they usually define absolute re-



(a) 1000 particles of radius 1 in (b) 2000 particles of radius 0.794 in (c) 4000 particles of radius 0.63 in

Figure 5: Tumbling mill filled with three particle sizes namely (a) 1 in, (b) 0.794 in and (c) 0.63 in using respectively 1000, 2000 and 4000 particles.



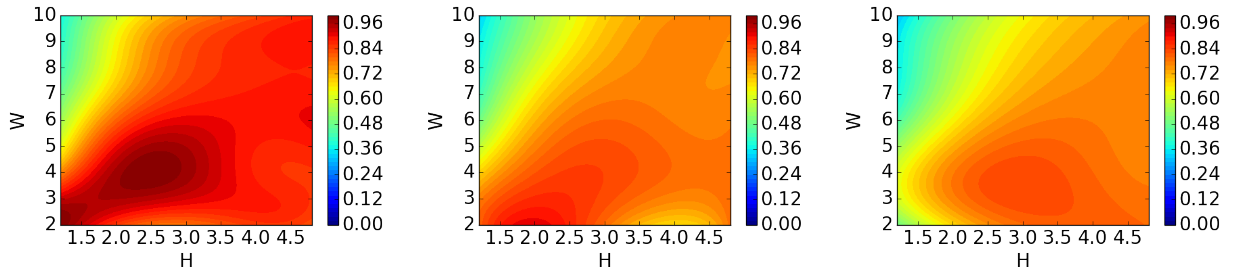
(a) 1000 particles of radius 1 in (b) 2000 particles of radius 0.794 in (c) 4000 particles of radius 0.63 in

Figure 6: Actual normalized power contours of the actual data, using the 20 full factorial design of experiment points, for the three particle sizes, namely (a) 1 in, (b) 0.794 in and (c) 0.63 in using respectively 1000, 2000 and 4000 particles.

strictions on the computed values. It is therefore important to note that starting with a too low numerical fidelity may be inadequate to save computational time when constraints are considered unless additional measures are taken which is part of our current investigation.

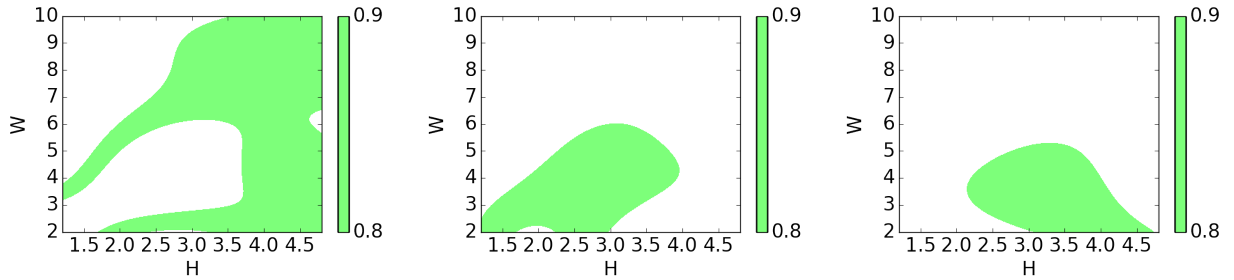
5 CONCLUSIONS

By combining lower fidelity models with high performance solution strategies such as efficient graphical processing unit (GPU) based discrete element modelling (DEM) enables us not to only do simulations faster but differently. This allows for the development of new and alternative paradigms of which interactive simulation and design was considered in this study, and essentially the paradigm for which the simulation environment BlazeDEM-



(a) 1000 particles of radius 1 in (b) 2000 particles of radius 0.794 in (c) 4000 particles of radius 0.63 in

Figure 7: RBF predicted normalized power contours of the actual data, using the 20 full factorial design of experiment points, for the three particle sizes, namely (a) 1 in, (b) 0.794 in and (c) 0.63 in using respectively 1000, 2000 and 4000 particles.



(a) 1000 particles of radius 1 in (b) 2000 particles of radius 0.794 in (c) 4000 particles of radius 0.63 in

Figure 8: All the designs that lies between the level sets 0.8 to 0.9 of the normalized power for the three particle sizes, namely (a) 1 in, (b) 0.794 in and (c) 0.63 in using respectively 1000, 2000 and 4000 particles.

GPU was developed. In addition to the design cycle, the role that this paradigm will play in education is invaluable as an in-house corporate training tool for young engineers to actively train and develop understanding for specific industrial processes. The initial results prove to be promising and warranting extensive research to be conducted in future.

Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

REFERENCES

[1] Clearly P. and Sawley M. DEM modelling of industrial granular flows: 3D case studies and the effect of particle shape on hopper discharge. *App. Math. Mod.* (2002)**26**:89–111.

[2] Govender N., Wilke D.N. and Kok S., Collision detection of convex polyhedra on the NVIDIA GPU architecture for the discrete element method, *App. Math. Comp.*. (2015)**267**:810–829.

[3] Boukouvala F., Muzzio F., Ierapetritou M. G. and Y. Gao, Reduced-order discrete element method modeling, *Chem. Eng. Sci.* (2013)**95**:12–26.

[4] *Enhanced Intel SpeedStep Technology for the Intel Pentium M Processor (White Paper)*. Intel Corporation. (2004) Retrieved 2013-12-21.

[5] Wilke D.N., Govender N., Rajamani R.K. and Pizette P. *Geometric design of tumbling mill lifter bars utilizing the discrete element method.* 12th World Congress on Structural and Multidisciplinary Optimization 5th - 9th, June 2017, Braunschweig, Germany.

[6] Robinson T. D., Eldred M. S. Willcox K. E. and Haimes R. Surrogate-Based Optimization Using Multifidelity Models with Variable Parameterization and Corrected Space Mapping. *AIAA J.*. (2008)**46**:2814–2822.

[7] Leifsson L. and Koziel S. Multi-fidelity design optimization of transonic airfoils using physics-based surrogate modeling and shape-preserving response prediction, *J. of Comp. Sci.*. (2010)**1**:98–106.

[8] Wilke, D.N., Kok S. and Groenwold A.A. Relaxed error control in shape optimization that utilizes remeshing. *Int. J. of Num. Meth. in Eng.* (2013)**94**:273–289.

[9] Allaire D. and Wilcox K. A mathematical and computational framework for multifidelity design and analysis with computer models. *Int. J. Uncert. Quant.*(2014)**(4)**:1–20.

[10] Yorke J.A. and Anderson Jr W.N. Predator-Prey Patterns (Volterra-Lotka equations). *Proceedings of the National Academy of Sciences.*(1973)**70**.

[11] Wang Y. *Performance and Power Optimization of GPU Architectures for General-purpose Computing.* PhD University of South Florida (2014). Accessed by Graduate theses and Dissertations `http://scholarcommons.usf.edu/etd/5325`

[12] Govender N., Wilke D.N., Kok S. and Els R. Development of a convex polyhedral discrete element simulation framework for NVIDIA Kepler based GPUs, *Journal of Computational and Applied Mathematics*).(2014)**270**:386–400.

[13] Wilke, D.N., Govender N., Rajamani R.K. and Pizette P. Geometric design of tumbling mill lifter bars utilizing the discrete element method. *12th World Congress of Structural and Multidisciplinary Optimisation.*(2017).