Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.jzus.zju.edu.cn; www.springerlink.com
E-mail: jzus@zju.edu.cn

Real-time energy management controller design for a hybrid excavator using reinforcement learning*

Qian ZHU†, Qing-feng WANG

(*State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China*)

†E-mail: 11125025@zju.edu.cn

Received Sept. 30, 2016; Revision accepted Mar. 28, 2017; Crosschecked Oct. 11, 2017

Abstract: Real-time energy management of a hybrid excavator is addressed using reinforcement learning (RL). Due to the computational complexity and need for *a priori* knowledge of the load cycles, a traditional optimal control method, like dynamic programming (DP), is not feasible for real-time control. Real-time controllers derived from traditional optimal control methods compute the solutions either in a cycle-dependent manner or far away from the optimal. An RL-based energy management controller is proposed to solve this problem. The simulation and experimental results demonstrate that the RL controller has a better performance than the widely used thermostat and equivalent consumption minimization strategy (ECMS) controllers. It also shows that the RL controller is cycle-independent. Pontryagin's minimum principle (PMP) is used to obtain the analytical solution of the energy management problem, and this can help to reduce the iteration time in the design process.

Key words: Energy management; Real time; Hybrid excavator; Reinforcement learning (RL); Pontryagin's minimum principle (PMP)

http://dx.doi.org/10.1631/jzus.A1600650 **CLC number:** TU621

1 Introduction

Excavators are heavy construction machines that are widely used in infrastructure construction. However, they consume a great deal of energy and involve severe pollution. Because of global warming and increasing oil prices, new technologies for excavators are required to improve fuel economy and reduce emissions. Hybrid technology, which has been first applied to vehicles successfully (Chan, 2007; Ehsani *et al*., 2007; Salmasi, 2007), is regarded as a promising solution to provide clean and efficient construction machinery. In recent years, it has been introduced to construction machinery especially excavators (Xiao *et al*., 2008; Lin *et al*., 2010).

Hybrid excavators can be divided into three categories: series, parallel, and compound. The parallel hybrid excavator (Fig. 1) has a relatively simple structure and comparable fuel economy compared to other types. The main ideas of the parallel type study can be extended to other types. Therefore, the parallel hybrid excavator is chosen as the object of this study. In this type, the engine, pump, and assist motor are directly connected to each other. The assist motor can work as a generator to charge the energy storage system or work as a motor to provide additional power for the drive train. An ultracapacitor is usually chosen as the energy storage unit because of its high specific power, fast dynamics, and long life.

A parallel hybrid excavator has more than one path of energy flow from the energy sources to the energy consumers. Energy management (Sciarretta and Guzzella, 2007), or energy flow control, is thus the key issue for effective operation. Performance characteristics, such as sustainability of charge, fuel economy, and emissions, strongly depend on the

* Project supported by the National Natural Science Foundation of China (No. 51475414) and the Science Fund for Creative Research Groups of National Natural Science Foundation of China (No. 51521064)

ORCID: Qian ZHU, http://orcid.org/0000-0002-7610-5103

energy management strategy. Optimal energy management can reduce fuel consumption significantly. Since the fuel cost of operating an excavator constitutes a large part of the total cost of ownership (TCO), energy management becomes far more important. Additionally, due to the complex structure, multiple work tasks, and aggressive load, energy management of a hybrid excavator is more challenging than that of a hybrid vehicle. Therefore, the optimal energy management problem should be carefully studied in order to obtain a good machine performance.

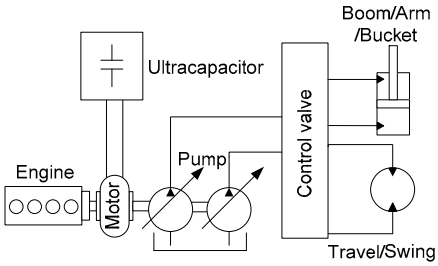

Fig. 1 Parallel hybrid excavator configuration

Energy management strategy for hybrid vehicles has been extensively studied (Salmasi, 2007; Sciarretta and Guzzella, 2007; Serrao *et al.*, 2011). The control methods can be classified into three categories, i.e., heuristic, optimal, and suboptimal methods.

Heuristic methods use rule-based controllers based on experience and engineering intuition (Jalil *et al.*, 1997; Salman *et al.*, 2000; Schouten *et al.*, 2002). These methods are easy to implement, since they are always simple and computationally efficient. However, heuristic methods do not guarantee global optimization. In fact, in most cases they are far away from the optimal solution.

Optimal methods are based on optimal control theories such as dynamic programming (DP) (Brahma *et al.*, 2000; Lin *et al.*, 2003) and Pontryagin's minimum principle (PMP) (Kim *et al.*, 2011; Kim and Rousseau, 2012). These methods can generate optimal solutions to the energy management problem, but require knowledge of the future power demand which is not known *a priori*. Optimal solutions are always time-dependent. Hence, they are non-causal strategies, which are not applicable in real world. Nevertheless, optimal methods can provide a benchmark for the performance of other strategies, and with proper modification can be used to develop online strategies.

Suboptimal methods are based on optimal methods of the previous category (Borhan *et al.*, 2009). They are modified optimal methods which are applicable to online use. Road prediction technology is always used in these methods (Zhang and Xiong, 2015) to make the optimal strategies applicable, but the results are not satisfactory. Additionally, accurate load prediction becomes far more challenging, when it comes to hybrid excavators, because of the particular working conditions. Even if the future load was predicted correctly, the computation complexity of the optimal optimization method is unacceptable for online application. Other suboptimal methods, extracting control laws from the optimal results, compute the solution in either a cycle-dependent manner or far away from the optimal. Another popular suboptimal method is the equivalent consumption minimization strategy (ECMS) based on PMP that can be used for real-time control (Musardo *et al.*, 2005; Serrao *et al.*, 2009). ECMS is only applicable for a specific drive cycle, since the equivalent factor is quite sensitive to the drive cycle.

For hybrid excavators, a few publications can be found on the energy management strategy, most of which are heuristic (Lin *et al.*, 2008; Xiao *et al.*, 2008) or suboptimal strategies (Kim *et al.*, 2012; 2016). Optimal energy management of this type has not yet been studied in depth. Since the structure and work load of hybrid excavators are very different from those of hybrid vehicles, optimal energy management strategy for hybrid vehicles cannot be directly applied here.

In this study a novel real-time optimal energy management strategy is proposed for the hybrid excavator based on RL. In this approach, the energy management problem is viewed as an infinite Markov process. The work load is regarded as an additional stochastic system state. The problem solution is constructed as a function of system state and work load. Thus, the optimal policy for deciding the assist motor output power becomes state-dependent with a feedback format, which is directly implementable in real time, rather than time-dependent (DP results). Load aggregation is introduced based on the

transition probability of the work load to apply the offline RL controller effectively. In order to have an insight into the optimal energy management strategy and also help to reduce the RL controller design time, the energy management problem is analytically solved through PMP. Finally, the simulation and experimental validation are undertaken.

2 System modeling

To develop an optimal controller based on a DP or RL algorithm, a proper system model or simulator of the hybrid excavator is needed. For the energy management problem, the hybrid excavator can be seen as a dynamic system with the only state variable the state of charge (SOC) of the ultracapacitor. Other fast dynamics, such as the engine and electric motor, are not considered. This quasi-static model is sufficient for energy management control. It is also favorable for implementing the optimal control methods because of low number of states and small computation time.

The engine model is based on a look-up table (fuel map shown in Fig. 2) which outputs fuel consumption as a function of the engine speed and torque. The fuel map is obtained from experimental data. Since the excavator operator always set the engine at a constant speed, the fuel consumption can be seen as a function of the engine speed and output power.

$$m_{\text{fuel}} = f(P_{\text{engine}}, \omega_{\text{engine}}), \tag{1}$$

where m_{fuel} is the fuel consumption, P_{engine} is the engine output power, and ω_{engine} is the engine speed.

The assist motor is modeled using the efficiency map shown in Fig. 3. The motor efficiency is a function of the motor torque and output power.

$$\eta_{\text{motor}} = f(P_{\text{motor}}, \omega_{\text{motor}}), \tag{2}$$

where η_{motor} is the efficiency of motor, P_{motor} is the motor output power, and ω_{motor} is the motor speed. Since the assist motor is directly coupled to the crankshaft in the parallel hybrid excavator, the motor speed and the engine speed are the same.

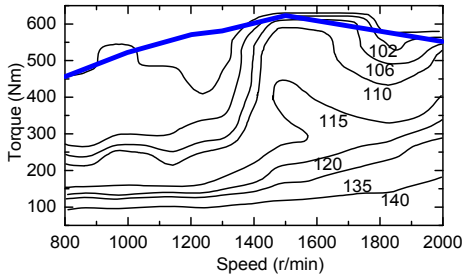

Fig. 2 Engine fuel consumption map

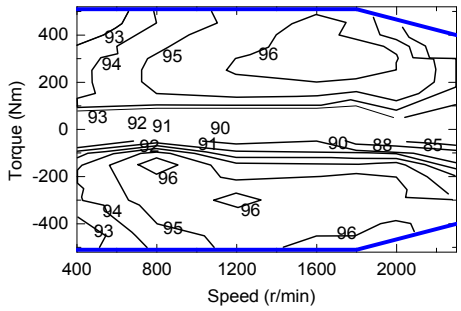

Fig. 3 Assist motor efficiency map

The ultracapacitor is modeled as an equivalent circuit comprising a voltage source in series with an effective internal resistance (Fig. 4).

$$P_{\text{c}} = (u_{\text{c}} i_{\text{c}} - R i_{\text{c}}^2) n, \tag{3}$$

$$\frac{\mathrm{d}u_{\text{c}}}{\mathrm{d}t} = \frac{i_{\text{c}}}{C}, \tag{4}$$

where P_{c} is the ultracapacitor power, i_{c} is the current in the ultracapacitor, u_{c} is the open circuit voltage of the ultracapacitor cell, R is the equivalent resistance, n is the number of capacitor cells, and C is the capacity. u_{c} is a function of the ultracapacitor SOC, while R is independent of the SOC. This differs from the case of a battery, in which the internal resistance is also a function of the SOC.

The SOC can be represented by the current voltage or energy of the ultracapacitor.

$$\text{SOC} = E_{\text{cur}} / E_{\text{max}}, \tag{5}$$

where E_{cur} is the current stored energy of the ultracapacitor, and E_{max} is the maximum energy that can be stored in the ultracapacitor.

The component limitations due to mechanical and electrical constraints are as follows:

$$\omega_{\text{engine}} \in \left[\omega_{\text{emin}}, \omega_{\text{emax}}\right], \tag{6}$$

$$P_{\text{engine}} \in \left[P_{\text{emin}}(\omega_{\text{engine}}), P_{\text{emax}}(\omega_{\text{engine}})\right], \tag{7}$$

$$\omega_{\text{motor}} \in \left[\omega_{\text{mmin}}, \omega_{\text{mmax}}\right], \tag{8}$$

$$P_{\text{motor}} \in \left[P_{\text{mmin}}(\omega_{\text{motor}}), P_{\text{mmax}}(\omega_{\text{motor}})\right], \tag{9}$$

where ω_{emin}, ω_{emax} and ω_{mmin}, ω_{mmax} are the minimum and maximum speeds of engine and motor, and P_{emin}, P_{emax} and P_{mmin}, P_{mmax} are the minimum and maximum powers of engine and motor, respectively.

The maximum power of the ultracapacitor is also limited by the maximum current. Here the maximum current is set to 400 A considering other component constraints and safety. The resulting maximum ultracapacitor power is much larger than the assist motor. Therefore, the limitation of the ultracapacitor power is not considered.

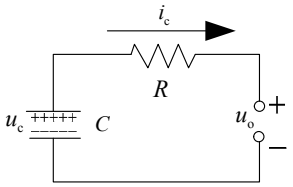

Fig. 4 Model of ultracapacitor (u_o is the output voltage of ultracapacitor)

3 Methods

The frameworks of the DP and RL algorithms are presented in this section. RL is employed to develop the real-time optimal controller. DP does not offer a real-time controller for the hybrid excavator energy management problem. However, it can serve as a baseline for other strategies. A rule-based thermostat strategy and the ECMS are also used for comparison. For more details of these two strategies, refer to (Johri and Filipi, 2009; Serrao *et al.*, 2009).

3.1 Dynamic programming

DP is applied to a discrete model of the form:

$$x_{k+1} = f(x_k, u_k). \tag{10}$$

The state variable x_k is the ultracapacitor SOC. The input variables u_k is the assist motor power. The cost function is

$$\min_{u(t)} J = \sum_{k=0}^{T} m_{\text{fuel}}, \tag{11}$$

where J is the total fuel consumption, and T is the length of the discrete load cycle. The problem is to find the optimal control u that minimizes the total fuel consumption during a given load cycle.

The DP algorithm can compute the optimal solution for a given load cycle. It is a multi-stage optimization method based on Bellman's optimality principle that performs the following equation backwards in time from N to 0.

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} \left[g_k(x_k, u_k) + J_{k+1}(x_{k+1})\right], \tag{12}$$

$$J_N(x_N) = g_N(x_N) = 0, \tag{13}$$

where N is the time length of the given load cycle, J is the optimal cost to go from state x_k at time k to the final time. U_k is the set of all feasible inputs u_k at state x_k. $g_k(x_k, u_k)$ is the cost of transitioning from state x_k to state x_{k+1} by applying an input u_k (Zhu *et al.*, 2016).

3.2 Reinforcement learning

RL is a direct adaptive optimal control method, primarily used for solving Markov and semi-Markov decision problems (Gosavi, 2003). For hybrid excavators, the load cycle can be modeled as a stochastic process with the load power being seen as a stochastic state. Such a stochastic process satisfies the Markov property, stating that the past is conditionally independent of the future given the present state of the process. Thus, the energy management problem in hybrid excavators is essentially a Markov decision process (MDP).

In an MDP, at each step the system is required to choose an action a from a finite set of actions according to a policy π. A policy is a state-action map mapping from the state set S to action set A. It defines the action to be chosen in every state. Then the state transits from state x_i to another state x_j with a probability $p(i, j)$. All the transition probabilities of states can be stored in a matrix called the transition probability matrix (TPM). For example, a TPM with three states is shown below. $\boldsymbol{P}(i, j)$ denotes the probability of transiting from state x_i to state x_j.

$$\boldsymbol{P}=\begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.5 & 0.1 \end{bmatrix}\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}. \qquad (14)$$
$$\quad x_1 \quad x_2 \quad x_3$$

The system receives an immediate reward after the state transitions. This is denoted by $r(i, j)$. The performance metric used to judge the performance of the policy is usually a function of the immediate rewards. There are two forms of the performance metric: the average reward and discounted reward. In this case, the action is the motor power P_{motor}, the states are SOC and the load power P_{load}, and the reward is the fuel consumption m_{fuel}.

Since the load power is seen as a state variable, the energy management problem becomes time-independent and does not have a final time defined. This is very different from the DP algorithm which computes a time-dependent solution for a given load cycle. Thus, the optimal solution obtained from the RL algorithm can be easily implemented online. The energy management problem for hybrid excavators is formulated as follows using a discounted reward over an infinite horizon:

$$\psi_i=\lim_{k\to\infty}E\left[\sum_{s=1}^{k}\gamma^{s-1}r\left(x_s,\pi(x_s),x_{s+1}\right)\right], \qquad (15)$$

where ψ_i is the discounted reward, and γ is the discount factor.

The MDP is solved by the famous Q-learning algorithm (Watkins and Dayan, 1992). RL can generally outperform heuristic and suboptimal methods, and at the same time the application of RL controller is much easier than DP. However, the design of an RL controller always needs a great deal of computation time and storage memory. These problems will be solved in Section 6.

It is important to understand that the system model is not necessary in the RL algorithm. We can implement the iteration in the real system choosing actions and observing the rewards. However, in this study we first develop the RL controller with the system model offline and then apply it online in order to reduce the iteration time.

4 Load aggregation

Direct online iteration of the RL algorithm is limited due to the large computation time. One possible approach is to first develop the RL controller offline using a simulator (system model) and then apply it to the real system while updating the policy matrix online.

The information on the work load is very important to the application of the RL offline controller. Different kinds of work load need different RL controllers. Therefore, load aggregation is needed so as to ensure an effective implementation of the offline controller. Load aggregation here means lumping several similar loads together. "Similar loads" means that the same controller can be applied in these loads. With aggregation of loads, one may apply the offline controllers to the real world according to the load categories. How to extract the load features and classify them is a difficult problem. Since the energy management problem is regarded as a Markov process in the RL algorithm and the optimal policy of a Markov process is determined by the system model and TPM model, load cycles of the same category should have a similar TPM.

Excavators are multifunction construction machines which can deal with different work tasks. Classifying the load cycles by the work tasks may be an effective method. This is usually the first thought when dealing with load aggregation. Load cycles of three typical work tasks (digging, grading, and lifting) are shown in Fig. 5. The transition probability of the load for each cycle is shown in Fig. 6. The load TPM can be obtained from the load cycle. In order to

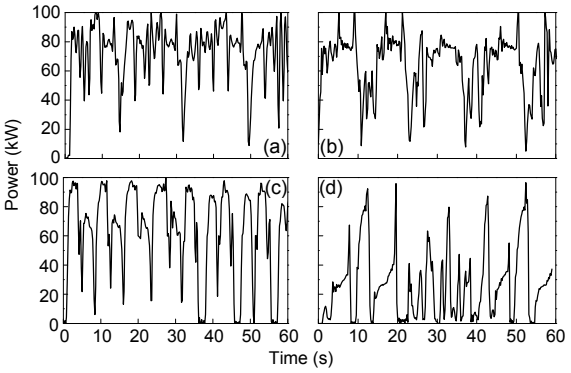

Fig. 5 Typical load cycles for excavators
(a) Digging 1 cycle; (b) Digging 2 cycle; (c) Grading cycle; (d) Lifting cycle

obtain the load TPM, the time and load variables of the load cycle need to be discretized. Here, the time discretization grid is set to 0.05 s and the load is discretized with a 3 kW interval. After discretization, the procedure is to find all the possible load states that state x_i can jump to in one step. Then count the occurrences when the load state transits from state x_i to state x_j and obtain $\boldsymbol{P}(i, j)$.

The load cycles were measured at 20 Hz and then filtered with a low pass filter of 5 Hz, which is sufficient to capture all the dynamics of the load cycles. The size of the load profile should be large enough to extract the load feature and obtain an accurate TPM. The load profile of each work task that we considered here has a length of more than 100 s.

From Fig. 6 we can see that the TPMs of different work tasks vary a lot, while the same work tasks have similar TPMs (the Nth load and (N+1)th load represent the present work load and the work load in the next time step, respectively). Therefore, load cycles of the same work task can be classified into one category. The distribution of TPMs can represent the load features. Centralized areas imply that the number of operation points is relatively small or that the load varies gently in this region. Decentralized areas imply a large number of operation points or an aggressive load in this region.

One thing to note is that the engine may operate at different speeds for the same work task according to the power of the load which makes the engine operate efficiently. The engine fuel consumption model also varies with speed. Therefore, load cycles of the same work task with different engine speeds are not aggregated to the same class.

5 Simulation results

In this section four control strategies (thermostat, ECMS, DP, and RL) are implemented in a simulator over the standard digging cycle (digging 1). DP serves as an optimal energy management strategy

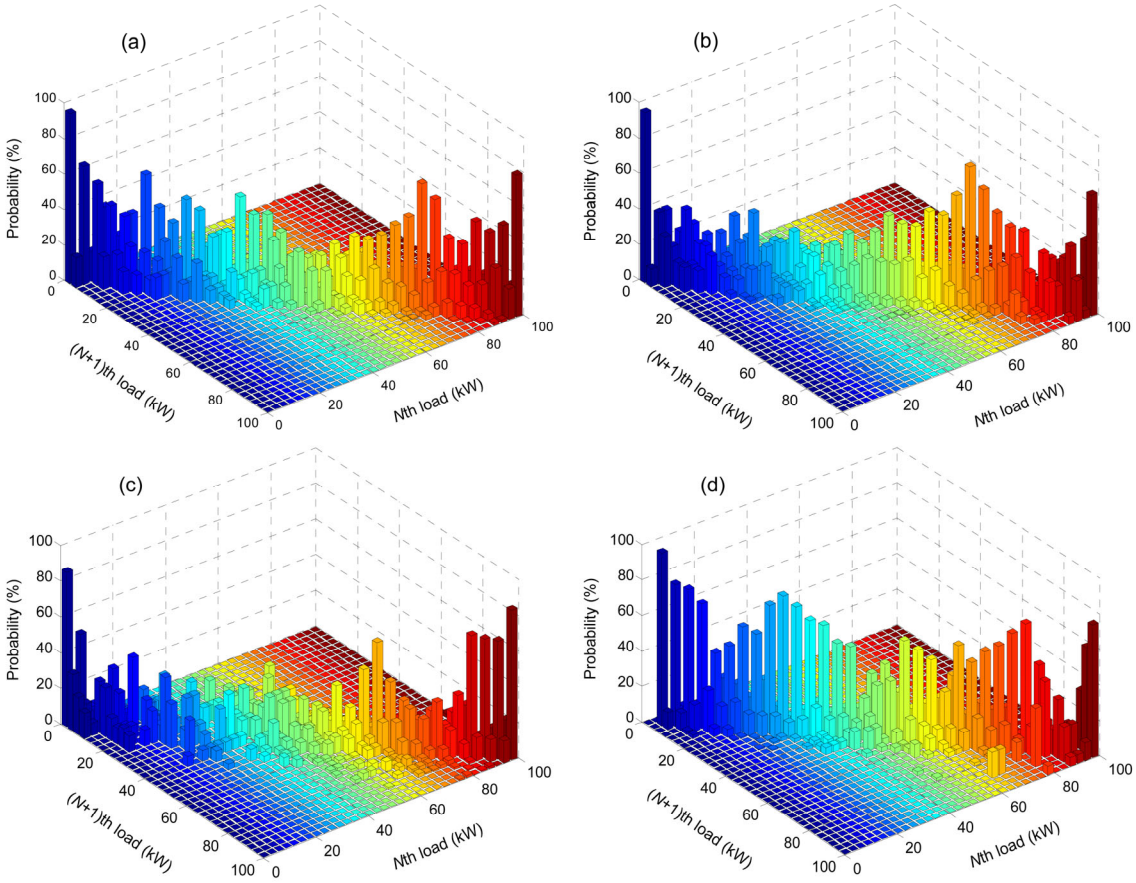

Fig. 6 TPM of different load cycles: (a) digging 1 cycle; (b) digging 2 cycle; (c) grading cycle; (d) lifting cycle

here for comparison. The corresponding fuel consumption for all these control strategies is presented in Table 1.

In order to make a comparison, the fuel consumption has been corrected according to the final SOC value with an average equivalent factor (Sciarretta *et al.*, 2004). It is observed that RL and ECMS 1 (with an optimal equivalent factor) can achieve almost the same result as DP, while thermostat and ECMS 2 (with a suboptimal equivalent factor) present a relatively poor performance. These results can be explained by the following analysis.

Table 1 Fuel consumption of the four strategies

Controller	Relative fuel consumption (%)
Traditional excavator	100
DP	91.3
Thermostat	96.5
ECMS 1 (optimal λ=3.39)	91.2
ECMS 2 (suboptimal λ=3.30)	94.1
RL	91.5

λ is the equivalent factor in ECMS controller

5.1 Comparison of thermostat and DP

Energy management results of the thermostat are shown in Figs. 7 and 8 together with the DP results. It can be seen that both the thermostat and DP strategies can regulate the engine operation points to the highly efficient region (high power region). Engine operation points of DP are more concentrated in the efficient region than those with the thermostat strategy. This is because a rule-based method is an engine-centric strategy, such that only the engine efficiency and instantaneous fuel economy are considered. However, the optimal method takes the overall efficiency and long-term fuel economy into consideration.

The motor power of the thermostat strategy is more dependent on the SOC rather than the load cycle. The thermostat strategy is therefore a general (cycle-independent) control strategy for hybrid excavator energy management control. The SOC of the thermostat strategy has a violent fluctuation and is more likely to violate the limitations, and this is not favorable for the ultracapacitor control.

5.2 Comparison of ECMS and DP

Energy management results of the ECMS 1 strategy are depicted in Figs. 9 and 10. The equivalent factor is chosen optimally through an iterative tuning method. It is observed that the SOC and motor power trajectory are very close to the optimal solution obtained from the DP. This is because ECMS is actually equivalent to the PMP, provided that the value of the equivalent factor is perfectly tuned.

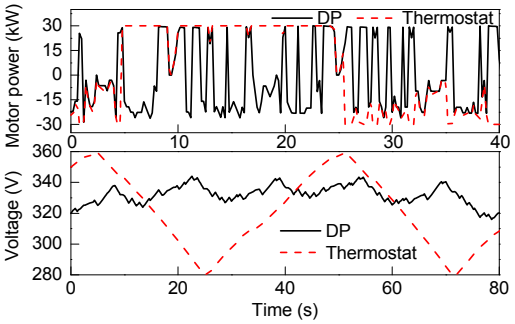

Fig. 7 Motor power and ultracapacitor voltage trajectories of thermostat and DP strategies

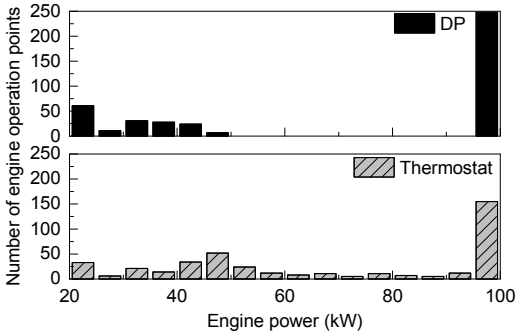

Fig. 8 Engine operation point distributions of thermostat and DP strategies

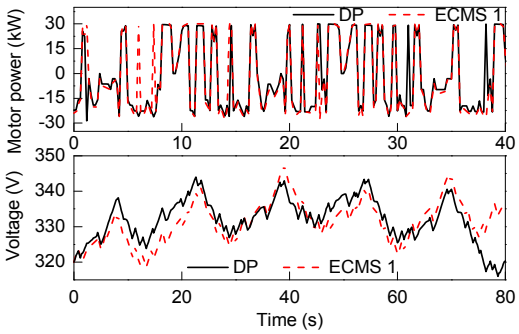

Fig. 9 Motor power and ultracapacitor voltage trajectories of ECMS 1 and DP strategies

However, the performance of ECMS is very sensitive to the equivalent factor. A small change of the factor may lead to a totally different result.

Energy management results of ECMS 2 with a suboptimal factor value are shown in Figs. 11 and 12. We can see that the motor behaves far away from the optimal trajectory in ECMS 2. The sustainability of charge is also no longer guaranteed.

From the simulation results, it can be concluded that ECMS can generate the optimal energy management only if the equivalent factor is properly tuned. ECMS is actually a cycle-dependent strategy since the value of the optimal equivalent factor depends on the load cycle. Therefore, online adaptation of the equivalent factor should be applied when implementing ECMS in the real world. An additional SOC management strategy is also needed to satisfy the charge-sustaining requirement. However, adaptation of the optimal equivalent factor is not that easy since the factor value is quite sensitive to the load cycle. Additionally, feedback control of the SOC could further deteriorate the performance.

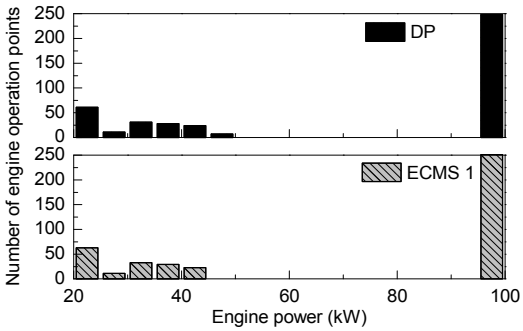

Fig. 10 Engine operation point distributions of ECMS 1 and DP strategies

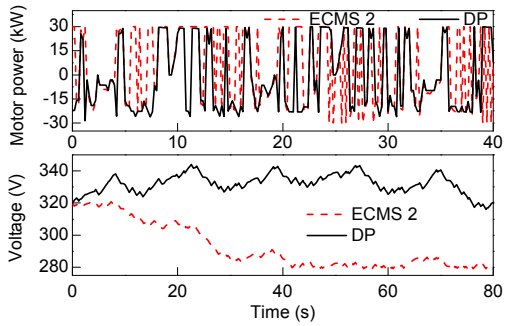

Fig. 11 Motor power and ultracapacitor voltage trajectories of ECMS 2 and DP strategies

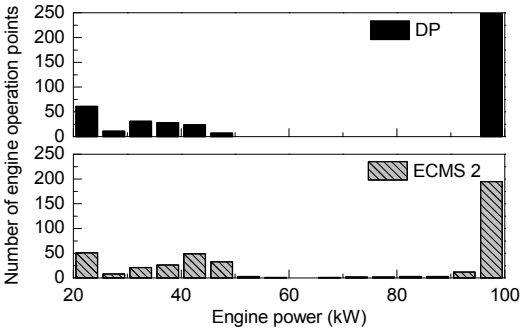

Fig. 12 Engine operation point distributions of ECMS 2 and DP strategies

5.3 Comparison of RL and DP

The RL policy is shown in Fig. 13. The motor power is a function of the SOC and load power. Energy management results of the RL strategy are shown in Figs. 14 and 15. It can be seen that the RL strategy can get almost the same performance as the optimal DP. This is because RL is actually an offshoot of DP (Gosavi, 2003). It is an optimal control algorithm which can be obtained from the Bellman's optimality principle. Thus, it is no surprise that the RL strategy can get a performance close to the optimal solution. Unlike DP, which is a non-causal algorithm, RL is an implementable energy management strategy. In the RL algorithm, the energy management problem is no longer a time-domain problem (as it is in the DP algorithm). The action (motor power) only depends on the current load state and system state (SOC). It is therefore a causal strategy, and can be applied online directly.

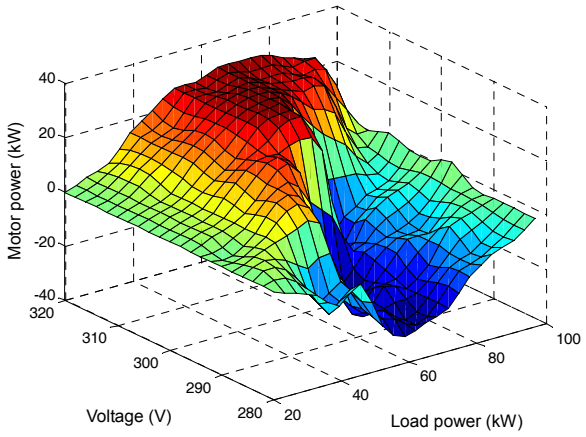

Fig. 13 Optimal RL policy

Compared to ECMS, RL is a very robust controller. It can perform well over different load cycles of the same work task, since they have the similar TPMs. In fact, the RL controller of a certain work task can also be implemented for other work tasks due to its state feedback property. However, performance may be degraded. It is a general energy management strategy like the thermostat strategy.

Previous results show that the RL strategy can outperform the heuristic and suboptimal strategies, and it is much easier to design a real-time controller compared to other optimal methods (DP and PMP). However, all these come at a price. Design of the RL controller needs a large number of iterations. Compared to a simplified DP in which the average time of one optimizing cycle is around 4 min, RL takes about 20 h to converge to an ideal result (with 2.0 GHz dual core CPU). This is because RL needs to compute the actions for all possible load and SOC state combinations, which needs a large number of iterations. The memory storage used to store the RL controller (Q matrix and policy matrix) is also larger than the rule-based controllers. These problems are addressed in the following section.

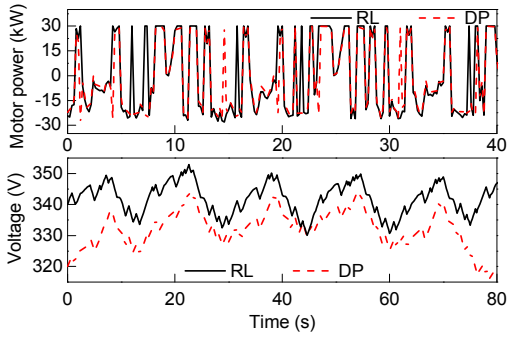

Fig. 14 Motor power and ultracapacitor voltage trajectories of RL and DP strategies

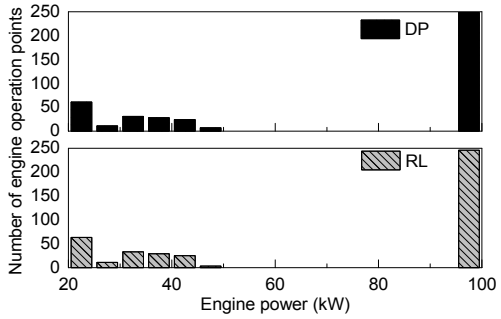

Fig. 15 Engine operation point distributions of RL and DP strategies

6 PMP-based analytical solution

PMP is used in this section to solve the energy management problem analytically. The analytical solutions can give an insight into how the assist motor behaves optimally with the varying load and state and also help to reduce the iteration time in RL.

The DP algorithm presented in the previous section is a numerical optimization method for a discrete-time system. It can only obtain numerical solution, from which it is hard to extract the control laws. An analytical DP algorithm for a continuous system needs to solve the Hamilton–Jacobi–Bellman (HJB) equation, which is a partial differential equation and is not always easy to solve. PMP is another analytical optimization algorithm, which can deal with optimal control problems with control constrains. It is much easier to implement PMP than DP, since it solves just differential equations.

The task of the optimal control problem is that, find a piecewise continuous control $u{:}[t_0, t_f]\rightarrow\Omega_u(t)$, such that the constraints

$$x(t_0)=x_0,\ \ x(t_f)=x_f,\ x(t)\in\Omega_x(t), \tag{16}$$

$$\dot{x}(t)=f(x(t),u(t),t),\ \ t\in[t_0,t_f] \tag{17}$$

are satisfied and such that the cost function

$$J(u)=\int_{t_0}^{t_f}L(x(t),u(t),t)\mathrm{d}t \tag{18}$$

is minimized. Note that the functions f and L are assumed to be at least once continuously differentiable with respect to all of their arguments (Geering, 2007). In Eqs. (16)–(18), u is the control variable, x is the state variable, t_0 is the initial time, t_f is the final time, and L is the cost function. For the energy management problem in hybrid excavators, the state variable x is the SOC. The control variable is the motor power P_{motor}. The cost function is the fuel consumption m_{fuel}. L can also be defined to include other performance characteristics like emissions and drivability. $\dot{x}(t)=f(x(t),u(t),t)$ is the system dynamics, and here only the ultracapacitor dynamics is considered. Ω are instantaneous constrains of the state and control variables.

Since this is a complex nonlinear problem with state and control constrains, it is hard to solve it directly. Some assumptions and approximations need to be made, in order to apply PMP.

Assumption 1 The state constrains are inactive for all $t\in[t_0, t_f]$. That is the SOC stays strictly within the boundaries at any given time. $x_{min}<x(t)<x_{max}$, $t\in[t_0, t_f]$.

Under assumption 1, the previously formulated problem becomes an optimal control problem with control constraints only, and is much easier to solve than problems with state constraints. Such an assumption is reasonable for the following reasons:

1. The repeated work cycle of excavators is always very short (no more than 10 s). Thus, the energy fluctuation of the ultracapacitor is small.

2. The energy management problem for hybrid excavators is a charge sustainability problem, in which the ultracapacitor operates in a narrow range. There is little likelihood of boundary violations. The optimal state trajectory from the DP results also validates this assumption.

Assumption 2 The efficiency of the electric system can be neglected.

This assumption ensures the continuity and differentiability of the cost function L, and this is a necessary condition of PMP. The internal resistance of the ultracapacitor is relatively small compared to that of the battery, and the voltage remains high throughout the work cycle. All these factors make the ultracapacitor a highly efficient unit. The PMP results also imply that the electric motor works in a highly efficient region as will be seen in the following section. On the other hand, the obtained PMP solution under this assumption is sufficient to analyze the optimal assist motor behavior or serves as an initial RL policy.

Approximation 1 The engine efficiency model $m_{fuel}=f(P_{engine}, \omega_{engine})$ is approximated by a quadratic polynomial equation of P_{engine}.

$$\begin{aligned} m_{fuel}(P_{engine}) &= m_{fuel}(P_{load}-P_{motor}) \\ &= a(P_{load}-P_{motor})^2+b(P_{load}-P_{motor})+c. \end{aligned} \tag{19}$$

Approximation 2 The load sequences are reordered in nondecreasing order and approximated by a fourth-order polynomial equation.

The order of the load sequence has little effect on the optimal solution as long as the state does not violate the boundaries over the whole time horizon. This has been validated in the DP results. One thing that needs to be noted is that the optimal analytical solution of PMP is a function of state and load rather than a function of time. Therefore, the order of load sequences does not affect obtaining the optimal solution. Nondecreasing order load sequences make the PMP much easier to solve as will be seen in the following section.

Approximation 2 together with approximation 1 and assumption 2 ensure continuity and differentiability of the cost function L. Under the assumption and approximations we redefine the optimal energy management problem as follows.

Find a piecewise continuous control $P_{motor}(t)$, such that the constraints

$$\dot{SOC}(t) = -P_{motor}(t), \text{ for all } t\in[t_0, t_f], \tag{20}$$

$$SOC(t_0) = SOC(t_f) = SOC_{target}, \tag{21}$$

$$SOC_{min} \le SOC(t) \le SOC_{max}, \tag{22}$$

$$P_{mmin} \le P_{motor}(t) \le P_{mmax}, \tag{23}$$

$$P_{emin} \le P_{engine}(t) \le P_{emax} \tag{24}$$

are satisfied and such that the cost functional

$$J(P_{motor}) = \int_{t_0}^{t_f} \dot{m}_{fuel}(P_{motor}(t), P_{load}(t))\,dt \tag{25}$$

is minimized.

All of the necessary conditions of PMP are formulated as follows. Using the cost function Eq. (25) leads to the Hamiltonian function:

$$\begin{aligned} &H(SOC(t), P_{motor}(t), P_{load}(t)) \\ &\quad= \dot{m}_{fuel}(P_{motor}(t), P_{load}(t)) + \lambda(t)\dot{SOC}(t) \\ &\quad= a(P_{load}-P_{motor})^2 + b(P_{load}-P_{motor}) + c - \lambda(t)P_{motor}. \end{aligned} \tag{26}$$

Pontryagin's necessary conditions for optimality are shown as follows. If $P^o_{motor}(t)$ is the optimal control, then the following conditions are satisfied:

$$\dot{SOC}^o(t) = -P^o_{motor}(t), \tag{27}$$

$$\dot{\lambda}^{\mathrm{o}}(t)=-\frac{\partial H}{\partial \mathrm{SOC}}=0, \tag{28}$$

$$\mathrm{SOC}^{\mathrm{o}}(t_0)=\mathrm{SOC}^{\mathrm{o}}(t_{\mathrm{f}})=\mathrm{SOC}_{\mathrm{target}}, \tag{29}$$

$$P_{\mathrm{mmin}}\le P^{\mathrm{o}}_{\mathrm{motor}}(t)\le P_{\mathrm{mmax}}, \tag{30}$$

$$P_{\mathrm{emin}}\le P^{\mathrm{o}}_{\mathrm{engine}}(t)\le P_{\mathrm{emax}}, \tag{31}$$

$$\begin{aligned}&H(\mathrm{SOC}^{\mathrm{o}}(t), P^{\mathrm{o}}_{\mathrm{motor}}(t), \lambda^{\mathrm{o}}(t), P_{\mathrm{load}}(t))\\ &\quad\le H(\mathrm{SOC}^{\mathrm{o}}(t), P_{\mathrm{motor}}(t), \lambda^{\mathrm{o}}(t), P_{\mathrm{load}}(t)).\end{aligned} \tag{32}$$

The differential equations for the co-state variables $\lambda^{\mathrm{o}}(t)$ imply that $\lambda^{\mathrm{o}}(t)\equiv\lambda_{\mathrm{o}}$ is constant. The Hamiltonian function is a quadratic polynomial function of P_{motor} with a negative coefficient a.

$$\frac{\partial H}{\partial P_{\mathrm{motor}}}=0\ \rightarrow\ P_{\mathrm{motor}}=P_{\mathrm{load}}+\frac{b+\lambda_{\mathrm{o}}}{2a}=P_{\mathrm{load}}+s_{\mathrm{o}}, \tag{33}$$

letting $s_{\mathrm{o}}=\dfrac{b+\lambda_{\mathrm{o}}}{2a}$.

Therefore, minimizing the Hamiltonian function yields the following control law:

$$P^{\mathrm{o}}_{\mathrm{motor}}(t)=\begin{cases}P_{\mathrm{mmin}}(t), & P_{\mathrm{load}}(t)+s_{\mathrm{o}}>\dfrac{P_{\mathrm{mmin}}(t)+P_{\mathrm{mmax}}(t)}{2},\\ P_{\mathrm{mmin}}(t)\ \text{or}\ P_{\mathrm{mmax}}(t), & P_{\mathrm{load}}(t)+s_{\mathrm{o}}=\dfrac{P_{\mathrm{mmin}}(t)+P_{\mathrm{mmax}}(t)}{2},\\ P_{\mathrm{mmax}}(t), & P_{\mathrm{load}}(t)+s_{\mathrm{o}}<\dfrac{P_{\mathrm{mmin}}(t)+P_{\mathrm{mmax}}(t)}{2}.\end{cases} \tag{34}$$

It is obvious that the motor always outputs the maximum or minimum power in the admissible set (somewhat like the bang-bang control). Under approximation 2, the load is a nondecreasing sequence. The solution $P^{\mathrm{o}}_{\mathrm{motor}}(t)$ in the time domain can be easily transferred to $P^{\mathrm{o}}_{\mathrm{motor}}(P_{\mathrm{load}})$ in a load domain, from which it is easy to extract real-time control laws. There exists a

$$P^{*}_{\mathrm{load}}=\frac{P_{\mathrm{mmin}}(P_{\mathrm{load}})+P_{\mathrm{mmax}}(P_{\mathrm{load}})}{2}-s_{\mathrm{o}} \tag{35}$$

satisfying

$$P^{\mathrm{o}}_{\mathrm{motor}}(P_{\mathrm{load}})=\begin{cases}P_{\mathrm{mmin}}(P_{\mathrm{load}}), & P_{\mathrm{load}}>P^{*}_{\mathrm{load}},\\ P_{\mathrm{mmin}}(P_{\mathrm{load}})\ \text{or}\ P_{\mathrm{mmax}}(P_{\mathrm{load}}), & P_{\mathrm{load}}=P^{*}_{\mathrm{load}},\\ P_{\mathrm{mmax}}(P_{\mathrm{load}}), & P_{\mathrm{load}}<P^{*}_{\mathrm{load}},\end{cases} \tag{36}$$

with P^{*}_{load} as the only unknown item, and it can be solved using the two state boundary conditions at the initial time and the final time:

$$\begin{aligned}\mathrm{SOC}^{\mathrm{o}}(t_{\mathrm{f}})-\mathrm{SOC}^{\mathrm{o}}(t_0)&=-\int_{t_0}^{t_{\mathrm{f}}}P_{\mathrm{motor}}(t)\,\mathrm{d}t\\ &=-\int_{P_{\mathrm{load}}(t_0)}^{P_{\mathrm{load}}(t_{\mathrm{f}})}P_{\mathrm{motor}}(P_{\mathrm{load}})\,\mathrm{d}P_{\mathrm{load}}=0.\end{aligned} \tag{37}$$

The optimal results of PMP are shown in Fig. 16 together with the DP results. We can see the two optimal results are quite similar. The small discrepancy may be caused by the assumptions and approximations. For instance, neglecting the electric system efficiency makes the assist motor work more favorably as a motor.

The somewhat bang-bang control obtained from PMP is due to the quadratic polynomial nature of the fuel consumption model. This approximation is accurate for engines with high efficiency in the high load region, which is common for diesel engines used in construction machinery. For engines with high efficiency in the medium load region, the approximation may be a third-order polynomial equation. In this case, the optimal control may locate where $\dfrac{\partial H}{\partial u}=0$. The optimal control is no longer a bang-bang control.

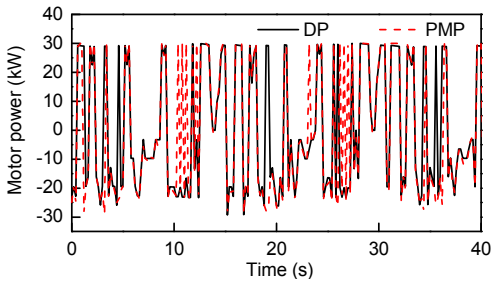

Fig. 16 Optimal motor power of DP and PMP

The optimal control policy from PMP does not consider the SOC. It assumes that the SOC stays within the boundaries, and this is validated by the SOC trajectory of the optimal DP results. However, in real-time application, boundary violation may occur when encountering unexpected aggressive loads. Therefore, a simple additional policy to deal with the SOC is needed. The SOC is divided into three sections: low, medium, high. The motor power is set to $P_{\mathrm{mmin}}(P_{\mathrm{load}})$ and $P_{\mathrm{mmax}}(P_{\mathrm{load}})$ at the low SOC and high SOC, respectively, while for a medium SOC the optimal control policy is applied.

The optimal control policy together with the SOC balance policy is shown in Fig. 17. We can see the optimal control policy from PMP is similar to that from the RL. It is sufficient to serve as an initial RL policy in the policy iteration approach to reduce the iteration time. It is found that the iteration time can be reduced to 1% of the previous design time after using the PMP policy.

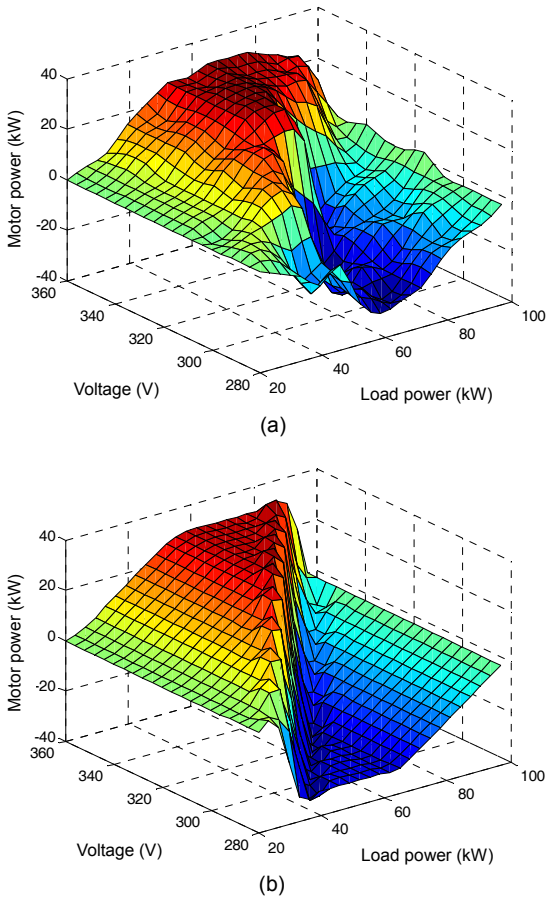

Fig. 17 Comparison of RL policy (a) and PMP policy (b)

The optimal control policy obtained from PMP can reduce the iteration time of the RL algorithm. However, there is still another difficulty when implementing RL. This is the curse of dimensionality. For our energy management problem, the MDP includes 100 SOC states, 25 load states, and 21 actions for each state. The Q matrix would contain 100×25×21 elements. It is difficult to store such a large matrix in the controller. State aggregation and function approximation technology are used to address this problem.

State aggregation combines states with similar characteristics together. With aggregation of states, we can obtain a relatively small Q matrix. From the RL and DP results we can see the optimal action is more sensitive to the load state than the SOC state. The load state is more important than the SOC state. Thus, SOC states with similar transition rewards can be aggregated to reduce the number of SOC states. The SOC can be discretized to 10 states. Thus, the Q matrix reduces to 10×25×21 elements.

Actor-critic algorithms are a class of approximate policy iteration techniques. They can further reduce the memory storage requirement. An actor-critic algorithm using neural networks is used here. The critic is the approximate value function, and the actor is the approximate policy. For more details refer to (Busoniu *et al.*, 2010).

7 Experimental results

The experimental validation of the previous proposed energy management strategies over the standard digging cycle is presented in this section. The test bench is a 20 t parallel hybrid excavator prototype (Fig. 18). The pump pressure is regulated by a proportional relief valve to simulate the standard digging cycle which ensures a fair comparison. Table 2 gives the specifications of the 20 t parallel hybrid excavator. The corresponding fuel consumption for the three online controllers is presented in Table 3.

Experimental results of the thermostat and RL controllers are shown in Figs. 19 and 20. We can see that compared to thermostat, RL can get much better fuel economy. RL can regulate more engine operation points to the highly efficient region (high torque

region). The SOC fluctuation range is also smaller compared to that of the thermostat.

Figs. 21 and 22 show the energy management results of the RL and ECMS controllers. It can be seen that the RL and ECMS controllers, which have a similar performance in simulation, vary a lot in experiment. RL can achieve better performance in terms of the fuel economy. RL also guarantees sustaining of the charge, while the SOC always violates the boundary in ECMS. When SOC reaches the boundary, the motor will be forced to work in a motor mode to reduce the SOC value (circle in Fig. 22). This will make the motor behave away from the optimal trajectory. This is because ECMS is essentially an open loop strategy and the equivalent factor is quite sensitive to the work load. Thus, a

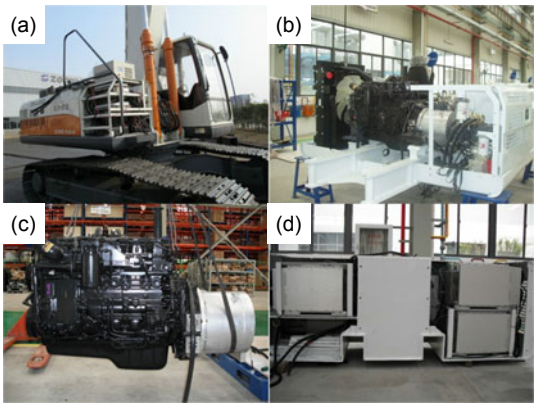

Fig. 18 Hybrid excavator prototype: (a) 20 t hybrid excavator; (b) power train; (c) engine and assist motor; (d) ultracapacitor

Table 2 Hybrid excavator prototype specifications

Component	Specification
Engine	Cummins QSB6.7 diesel, 115 kW@2000 r/min, 615 Nm@1500 r/min
Assist motor	Permanent magnet, 100 kW@2000 r/min, 500 Nm@1800 r/min
Hydraulic pump	Kawasaki K3V112DT, 2×112 ml/r, 34.3 MPa rated pressure
Ultracapacitor	Bainacap SCPC101796, 100 V rated voltage, 79 F rated capacity

Table 3 Fuel consumption of the three online controllers

Controller	Relative fuel consumption (%)
Traditional excavator	100
Thermostat	97.3
ECMS 1 (optimal λ=3.39)	94.5
RL	93.1

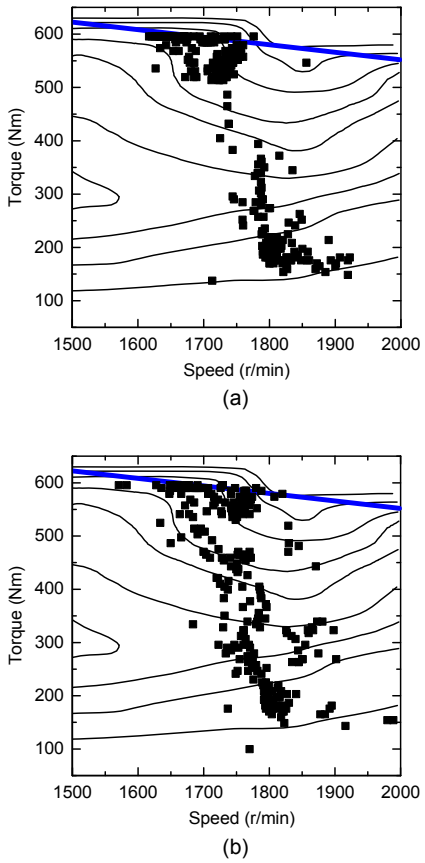

Fig. 19 Engine operation points: (a) RL; (b) thermostat

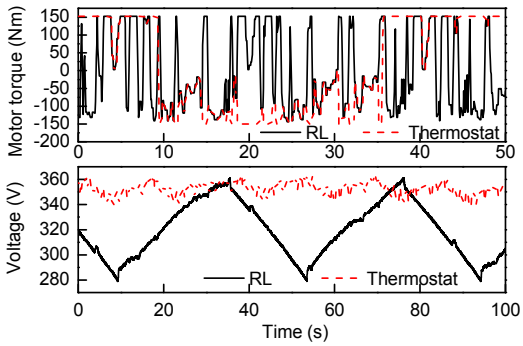

Fig. 20 Comparison of experimental results of thermostat and RL strategies

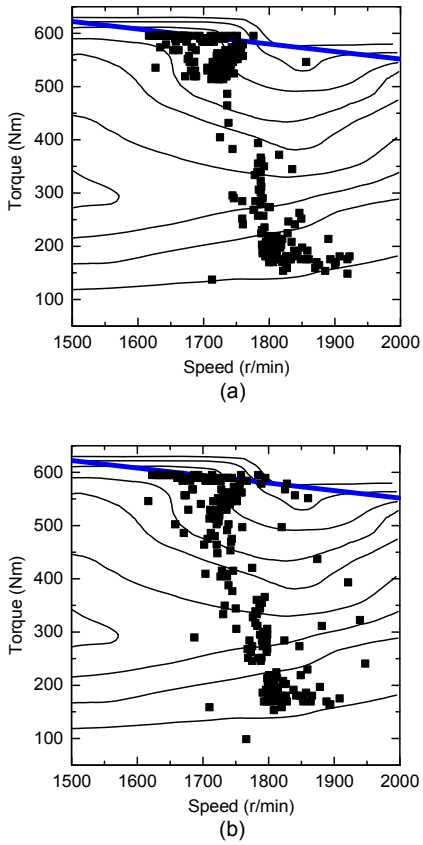

Fig. 21 Engine operation points: (a) RL; (b) ECMS

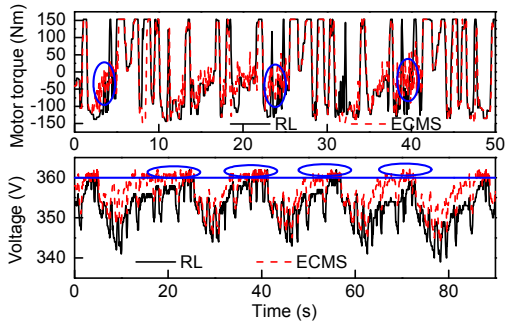

Fig. 22 Comparison of experimental results of ECMS and RL strategies

small error in the load cycle and system model may degrade the performance.

Online adaption of the equivalent factor is rather difficult compared to simulation. Fig. 23 shows the SOC trajectories of two different equivalent factors (ECMS 1, λ=3.37; ECMS 2, λ=3.41). A small change of the factor value can lead to a big difference in SOC behavior.

Fig. 24 shows the performance of the RL and ECMS (λ=3.37) controllers over another digging cycle (digging 2). It is observed that the RL controller is cycle-independent while the ECMS controller is cycle-dependent. The ECMS (λ=3.37) controller, which performs well in digging 1 cycle, fails to guarantee charge sustaining in the digging 2 cycle.

The ECMS strategy assumes that the equivalent factor is a constant and uses only one factor to deal with the energy management problem while RL uses a policy matrix. Therefore, it is no surprise that RL strategy can outperform the ECMS. Although the cycle-independent property of the RL controller only refers to cycles with the same work task, it is sufficient for excavator application since excavators repeat the typical work tasks most of the time.

Note that although all the energy management controllers are designed offline, the RL controller can be calibrated online. The policy matrix can also be updated online. In this experiment only a few iterations of calibration are carried to reduce the modeling error, and online update is not applied due to the limited computation ability and memory space of the control unit. The experimental results also demonstrate that an offline RL controller with online calibration is enough for application.

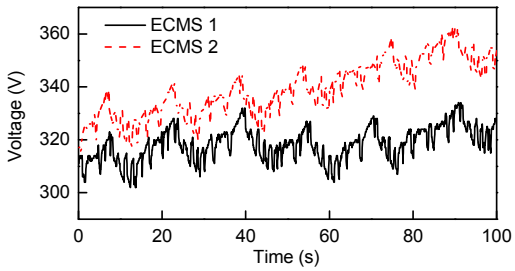

Fig. 23 Comparison of ultracapacitor voltages of ECMS 1 and ECMS 2

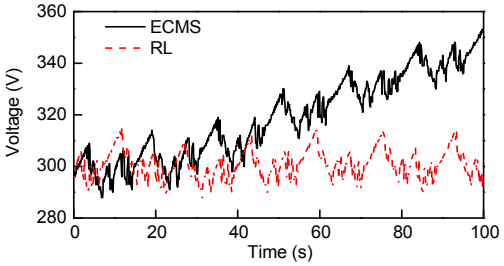

Fig. 24 Comparison of ultracapacitor voltages of ECMS and RL over digging 2 cycle

8 Conclusions

In this paper RL is used to develop a real-time energy management controller for the hybrid excavator. In this approach, the energy management problem is regarded as an infinite horizon MDP with the load power being seen as an additional stochastic state. Compared to another optimal control method (DP), RL can generate a time-independent controller which can be directly implemented in the real world. Simulation and experimental results show that the RL controller outperforms the widely used heuristic and sub-optimal controllers (thermostat and ECMS). It also demonstrates that the RL controller is applicable for different kinds of load cycles because of its closed-loop nature. Additionally, the analytical PMP solution is used as the initial policy when designing the RL controller in order to reduce the offline iteration time.

References

Borhan, H.A., Vahidi, A., Phillips, A.M., *et al*., 2009. Predictive energy management of a power-split hybrid electric vehicle. American Control Conference, p.3970-3976. http://dx.doi.org/10.1109/ACC.2009.5160451

Brahma, A., Guezennec, Y., Rizzoni, G., 2000. Optimal energy management in series hybrid electric vehicles. American Control Conference, p.60-64.

Busoniu, L., Babuska, R., Schutter, D., *et al*., 2010. Reinforcement Learning and Dynamic Programming Using Function Approximators. CRC Press, Boca Raton, USA, p.101-113. http://dx.doi.org/10.1201/9781439821091

Chan, C.C., 2007. The state of the art of electric, hybrid, and fuel cell vehicles. *Proceedings of the IEEE*, **95**(4): 704-718. http://dx.doi.org/10.1109/JPROC.2007.892489

Ehsani, M., Gao, Y., Miller, J.M., 2007. Hybrid electric vehicles: architecture and motor drives. *Proceedings of the IEEE*, **95**(4):719-728. http://dx.doi.org/10.1109/JPROC.2007.892492

Geering, H.P., 2007. Optimal Control with Engineering Applications. Springer, New York, USA, p.3-5.

Gosavi, A., 2003. Simulation-based Optimization. Springer, New York, USA, p.197-211. http://dx.doi.org/10.1007/978-1-4757-3766-0

Jalil, N., Kheir, N.A., Salman, M., 1997. A rule-based energy management strategy for a series hybrid vehicle. American Control Conference, p.689-693.

Johri, R., Filipi, Z., 2009. Low-cost pathway to ultra efficient city car: series hydraulic hybrid system with optimized supervisory control. *SAE International Journal of Engines*, **2**(2):505-520. http://dx.doi.org/10.4271/2009-24-0065

Kim, H., Choi, J., Yi, K., 2012. Development of supervisory control strategy for optimized fuel consumption of the compound hybrid excavator. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **226**(12):1652-1666.

Kim, H., Yoo, S., Cho, S., *et al*., 2016. Hybrid control algorithm for fuel consumption of a compound hybrid excavator. *Automation in Construction*, **68**:1-10. http://dx.doi.org/10.1016/j.autcon.2016.03.017

Kim, N., Rousseau, A., 2012. Sufficient conditions of optimal control based on Pontryagin's minimum principle for use in hybrid electric vehicles. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **226**(9):1160-1170.

Kim, N., Cha, S., Peng, H., 2011. Optimal control of hybrid electric vehicles based on Pontryagin's minimum principle. *IEEE Transactions on Control Systems Technology*, **19**(5):1279-1287. http://dx.doi.org/10.1109/TCST.2010.2061232

Lin, C.C., Peng, H., Grizzle, J.W., *et al*., 2003. Power management strategy for a parallel hybrid electric truck. *IEEE Transactions on Control Systems Technology*, **11**(6):839-849. http://dx.doi.org/10.1109/TCST.2003.815606

Lin, T., Wang, Q., Hu, B., *et al*., 2010. Development of hybrid powered hydraulic construction machinery. *Automation in Construction*, **19**(1):11-19. http://dx.doi.org/10.1016/j.autcon.2009.09.005

Lin, X., Pan, S.X., Wang, D.Y., 2008. Dynamic simulation and optimal control strategy for a parallel hybrid hydraulic excavator. *Journal of Zhejiang University-SCIENCE A*, **9**(5):624-632. http://dx.doi.org/10.1631/jzus.A071552

Musardo, C., Rizzoni, G., Guezennec, Y., *et al*., 2005. A-ECMS: an adaptive algorithm for hybrid electric vehicle energy management. *European Journal of Control*, **11**(4-5):509-524. http://dx.doi.org/10.3166/ejc.11.509-524

Salman, M., Schouten, N.J., Kheir, N.A., 2000. Control strategies for parallel hybrid vehicles. American Control Conference, p.524-528.

Salmasi, F.R., 2007. Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends. *IEEE Transactions on Vehicular Technology*, **56**(5):2393-2404. http://dx.doi.org/10.1109/TVT.2007.899933

Schouten, N.J., Salman, M.A., Kheir, N.A., 2002. Fuzzy logic control for parallel hybrid vehicles. *IEEE Transactions on Control Systems Technology*, **10**(3):460-468. http://dx.doi.org/10.1109/87.998036

Sciarretta, A., Guzzella, L., 2007. Control of hybrid electric vehicles. *IEEE Control Systems*, **27**(2):60-70.

http://dx.doi.org/10.1109/MCS.2007.338280

Sciarretta, A., Back, M., Guzzella, L., 2004. Optimal control of parallel hybrid electric vehicles. *IEEE Transactions on Control Systems Technology*, **12**(3):352-363. http://dx.doi.org/10.1109/TCST.2004.824312

Serrao, L., Onori, S., Rizzoni, G., 2009. ECMS as a realization of Pontryagin's minimum principle for HEV control. American Control Conference, p.3964-3969. http://dx.doi.org/10.1109/ACC.2009.5160628

Serrao, L., Onori, S., Rizzoni, G., 2011. A comparative analysis of energy management strategies for hybrid electric vehicles. *Journal of Dynamic Systems, Measurement, and Control*, **133**(3):031012. http://dx.doi.org/10.1115/1.4003267

Watkins, C., Dayan, P., 1992. Q-learning. *Machine Learning*, **8**(3-4):279-292. http://dx.doi.org/10.1007/BF00992698

Xiao, Q., Wang, Q., Zhang, Y., 2008. Control strategies of power system in hybrid hydraulic excavator. *Automation in Construction*, **17**(4):361-367. http://dx.doi.org/10.1016/j.autcon.2007.05.014

Zhang, S., Xiong, R., 2015. Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming. *Applied Energy*, **155**:68-78. http://dx.doi.org/10.1016/j.apenergy.2015.06.003

Zhu, Q., Wang, Q., Chen, Q., 2016. Component sizing for compound hybrid excavators. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **230**(7):969-982.

中文概要

题　目: 基于强化学习的混合动力挖掘机实时能量管理控制器设计

目　的: 混合动力挖掘机的能量管理策略直接影响着系统的燃油经济性。本文旨在通过研究混合动力挖掘机能量管理系统，得到最优能量管理策略，并开发实时能量管理控制器，降低系统的燃油消耗。

创新点: 1. 通过强化学习算法，设计时间无关的实时能量管理控制器；2. 通过极大值原理求得最优能量管理问题的解析解，并用来辅助实时能量管理控制器设计。

方　法: 1. 建立负载的马尔科夫模型，运用强化学习算法，得到实时能量管理控制器；2. 运用极大值原理，求得最优能量管理问题的解析解，并将其作为初始能量管理策略；3. 通过仿真模拟和实验研究，验证所设计的实时能量控制器的性能。

结　论: 1. 基于强化学习的能量管理控制器是一个可以在线应用的与时间无关的实时能量管理控制器；2. 基于强化学习的能量管理控制器优于广泛使用的恒温控制器和等效消耗最小化策略控制器；3. 基于强化学习的能量管理控制器由于其闭环特性可适用于不同类型的作业工况。

关键词: 能量管理；实时性；混合动力挖掘机；强化学习；极大值原理