# Heat Kernel Estimates for Jump Processes of Mixed Types on Metric Measure Spaces

**Zhen-Qing Chen*** and **Takashi Kumagai** †

Dedicated to Professor Masatoshi Fukushima on the occasion of his 70th birthday

(November 29, 2006)

**Abstract**

In this paper, we investigate symmetric jump-type processes on a class of metric measure spaces with jumping intensities comparable to radially symmetric functions on the spaces. The class of metric measure spaces includes the Alfors $d$-regular sets, which is a class of fractal sets that contains geometrically self-similar sets. A typical example of our jump-type processes is the symmetric jump process with jumping intensity

$$e^{-c_0(x,y)|x-y|}\int_{\alpha_1}^{\alpha_2}\frac{c(\alpha,x,y)}{|x-y|^{d+\alpha}}\,\nu(d\alpha)$$

where $\nu$ is a probability measure on $[\alpha_1,\alpha_2]\subset(0,2)$, $c(\alpha,x,y)$ is a jointly measurable function that is symmetric in $(x,y)$ and is bounded between two positive constants, and $c_0(x,y)$ is a jointly measurable function that is symmetric in $(x,y)$ and is bounded between $\gamma_1$ and $\gamma_2$, where either $\gamma_2\geq\gamma_1>0$ or $\gamma_1=\gamma_2=0$. This example contains mixed symmetric stable processes on $\mathbb{R}^n$ as well as mixed relativistic symmetric stable processes on $\mathbb{R}^n$. We establish parabolic Harnack principle and derive sharp two-sided heat kernel estimate for such jump-type processes.

## Contents



*The research of this author is supported in part by NSF Grants DMS-0303310 and DMS-06000206.

†The research of this author is supported in part by the Grant-in-Aid for Scientific Research (B) 18340027.

# 1 Introduction

Markov process with jumps or non-local Markovian operators has received much attention recently due to its importance in theory as well as in applications. Two-sided heat kernel estimates and parabolic inequalities have a long history in the theory of partial differential equations (PDEs). There are many beautiful results in these areas, which played fundamental role in the study of PDEs. On the contrary, their study for general stable-like processes (or fractional Laplacian-like operators) in $\mathbb{R}^d$ have only been studied very recently. In [18], Kolokoltsov obtained two-sided heat kernel estimate for certain stable-like processes in $\mathbb{R}^d$. Bass and Levin [6] used a completely different approach to obtain a similar estimate and a parabolic Harnack inequality for discrete time Markov chain on $\mathbb{Z}^d$ where the conductance between $x$ and $y$ is comparable to $|x-y|^{-(d+\alpha)}$ for $0 < \alpha < 2$. In [11], a two-sided heat kernel estimate and a scale-invariant parabolic Harnack inequality for symmetric $\alpha$-stable-like operators (of fixed order) on $d$-sets are obtained in [11]. (See [16] for some extensions.) So far the two-sided heat kernel estimate for non-local operators has been restricted to fixed order only. See [1] for some result on parabolic Harnack inequality and heat kernel estimate for non-local operators of variable order. See also [3, 4, 5, 7, 8, 21, 22] for related Harnack inequalities, Hölder continuity of harmonic functions and the Feller property for non-local operators.

This paper is concerned with heat kernel estimates and a scale-invariant parabolic Harnack inequality for a class of non-local symmetric operators of *variable order* on metric-measure spaces, which in particular include Alfors $d$-regular sets in Euclidean spaces. A prototype of the model consider in this paper when the state space is $\mathbb{R}^n$ is the following. Let

$$\begin{aligned}\mathcal{E}(f,f) &= \int_{\mathbb{R}^n\times\mathbb{R}^n} (f(x)-f(y))^2\, J(x,y)\, dxdy \\ \mathcal{F} &= \{f\in L^2(\mathbb{R}^n,dx):\ \mathcal{E}(f,f)<\infty\},\end{aligned}$$

where $J(x,y)$ is a symmetric kernel given by

$$J(x,y) := e^{-c_0(x,y)|x-y|}\int_{\alpha_1}^{\alpha_2} \frac{c(\alpha,x,y)}{|x-y|^{d+\alpha}}\,\nu(d\alpha)$$

Here $\nu$ is a probability measure on $[\alpha_1,\alpha_2]\subset(0,2)$, $c(\alpha,x,y)$ is a jointly measurable function that is symmetric in $(x,y)$ and is bounded between two positive constants, and $c_0(x,y)$ is a jointly measurable function that is symmetric in $(x,y)$ and is either bounded between two positive constants or is identically zero. It is not difficult to show that $(\mathcal{E},\mathcal{F})$ is a regular Dirichlet form on $\mathbb{R}^n$ and so there is a Hunt process $Y$ associated with it. The main result of this paper implies that $Y$ has a jointly continuous transition density function $p(t,x,y)$ on $(0,\infty)\times\mathbb{R}^n\times\mathbb{R}^n$, and consequently $Y$ can be refined to start from *every point* in $\mathbb{R}^n$. Moreover, a two-sided heat kernel estimate for $p(t,x,y)$ and a parabolic Harnack inequality are obtained. Note that this example includes the following particular cases:

(i) $Y$ is the independent sum of isotropically symmetric $\alpha_i$-stable processes on $\mathbb{R}^n$ with $\{\alpha_1,\cdots,\alpha_k\}\subset(0,2)$. This case corresponds to $c(\alpha,x,y)\equiv c(\alpha)$, $c_0(x,y)\equiv 0$ and $\nu=\frac{1}{k}\sum\delta_{\{\alpha_i\}}$;

(ii) $Y$ is a relativistic $\alpha$-stable process on $\mathbb{R}^n$, see Example 2.4 below;

(iii) $Y$ is the independent sum of relativistic $\alpha_i$-stable processes on $\mathbb{R}^n$ with $\{\alpha_1,\cdots,\alpha_k\}\subset(0,2)$.

In fact, in this paper we will consider a more general metric-measure space $F$ rather than just $\mathbb{R}^n$. In the following, if $f$ and $g$ are two functions defined on a set $D$, $f\asymp g$ means that there exists $C>0$ such that $C^{-1}f(x)\le g(x)\le C\,f(x)$ for all $x\in D$.

Let $(F,\rho,\mu)$ be a locally compact separable metric space with metric $\rho$ and a Radon measure $\mu$ having full support on $F$. For $x\in F$ and $r>0$, let $B(x,r)$ denote the open ball centered at $x$ with radius $r$. We assume that there is a point $x_0\in F$, a constant $\kappa\in(0,1]$, and an increasing sequence $r_n\to\infty$ so that for every $n\ge1$, $0<r<1$, and $x\in\overline{B(x_0,r_n)}$,

$$\text{there is some ball } B(y,\kappa r)\subset B(x,r)\cap\overline{B(x_0,r_n)}. \tag{1.1}$$

The above condition is satisfied when $F$ is bounded since in this case we can take $r_n$ strictly larger than the diameter of $F$ and take $\kappa=1$. The above condition is also satisfied when the metric $\rho$ on

$F$ is geodesic in the sense that for every $x, y \in F$, there is a is a continuous map $\gamma : [0,1] \to F$ with $\gamma(0) = x$ and $\gamma(1) = y$ such that

$$\rho(\gamma(s), \gamma(t)) = |t - s|\rho(x, y) \qquad \text{for every } t, s \in [0,1].$$

(Such a path $\{\gamma(t) : 0 \le t \le 1\}$ is called a geodesic path connecting $x$ and $y$.) This is because when $(F, \rho)$ is geodesic, fix some $x_0 \in F$, then for every integer $n \ge 1$, $r \in (0,1)$ and $x \in \overline{B(x_0, n)}$, there is a geodesic $\gamma$ connecting $x_0$ to $x$. Let $y$ be the point on $\gamma$ so that $\rho(x_0, y) = 1 - r/2$. Then $B(y, r/2) \subset B(x, r) \cap B(x_0, n)$ and so condition (1.1) is satisfied with $\kappa = 1/2$ and $r_n = n$. We point out that a geodesic space $(F, \rho)$ is pathwise connected while condition (1.1) allows for disconnected space $F$.

We also assume throughout this paper, unless otherwise specified, that $\mu(F) = \infty$, every ball $B(x, r)$ is relatively compact in $F$ and the volume doubling property: there exists $c > 0$ such that

$$\mu(B(x, 2r)) \le c\mu(B(x, r)) \qquad \text{for every } x \in F \text{ and } r \ge 0. \tag{$VD$}$$

Let $\phi$ be a strictly increasing continuous function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\phi(0) = 0$, and $\phi(1) = 1$ such that there exist constants $c > 0$ and $\beta_1 > 0$ such that

$$\frac{\phi(R)}{\phi(r)} \ge c\left(\frac{R}{r}\right)^{\beta_1} \qquad \text{for every } 0 < r < R < \infty \tag{1.2}$$

and that

$$\int_0^r \frac{s}{\phi(s)}\, ds < \infty \qquad \text{for every } r > 0. \tag{1.3}$$

Let $d$ be the diagonal of $F \times F$ and $J$ be a symmetric measurable function on $F \times F \setminus d$ such that for every $(x, y) \in F \times F \setminus d$,

$$\frac{c_1}{\mu(B(x, \rho(x, y))\phi(c_2\rho(x, y))} \le J(x, y) \le \frac{c_3}{\mu(B(x, \rho(x, y))\phi(c_4\rho(x, y))}. \tag{1.4}$$

When no confusion occurs, we will simply denote this as $J(x, y) \asymp \frac{1}{\mu(B(x, \rho(x,y))\phi(\rho(x,y))}$.

For $u \in L^2(F, \mu)$, define

$$\mathcal{E}(u) := \mathcal{E}(u, u) := \int_{F \times F} (u(x) - u(y))^2 J(x, y)\mu(dx)\mu(dy) \tag{1.5}$$

and for $\beta > 0$,

$$\mathcal{E}_\beta(u) := \mathcal{E}_\beta(u, u) := \mathcal{E}(u, u) + \beta \int_F u(x)^2 \mu(dx).$$

Let $C_c(F)$ denote the space of continuous functions with compact support in $F$, equipped with the uniform topology. Define

$$\mathcal{D}(\mathcal{E}) := \{f \in C_c(F) : \mathcal{E}(f) < \infty\}. \tag{1.6}$$

It will be shown in Proposition 2.2 below that $(\mathcal{E}, \mathcal{F})$ is a regular Dirichlet form on $L^2(F, \mu)$, where $\mathcal{F} := \overline{\mathcal{D}(\mathcal{E})}^{\mathcal{E}_1}$. So there is a Hunt process $Y$ associated with it on $F$, starting from quasi-every point in $F$ (see [14]). We denote the exceptional set as $\mathcal{N}_0$. Our goal is to derive two-sided estimate on the transition density function of $Y$ and to show that the process can be refined to start from every point in $F$. To this end, instead of (VD), we need a uniform volume doubling assumption: there exists strictly increasing function $V : \mathbb{R}_+ \to \mathbb{R}_+$ such that there is $c > 1$ so that

$$V(0) = 0 \qquad \text{and} \qquad V(2r) \le cV(r) \qquad \text{for every } r > 0 \tag{1.7}$$

and

$$\mu(B(x, r)) \asymp V(r) \qquad \text{for every } x \in F \text{ and } r > 0. \tag{1.8}$$

Under condition (1.8), condition (1.4) becomes

$$J(x, y) \asymp \frac{1}{V(\rho(x, y))\phi(\rho(x, y))} \qquad \text{for every } (x, y) \in F \times F \setminus d. \tag{1.9}$$

Note that (1.7) is equivalent to the following: there exists $c > 0$ and $d > 0$ such that

$$V(0) = 0 \qquad \text{and} \qquad \frac{V(R)}{V(r)} \le c\Big(\frac{R}{r}\Big)^d \qquad \text{for all } 0 < r \le R. \tag{1.10}$$

For most part of this paper, we assume the following stronger conditions than (1.2), (1.3) and (1.10): there exist constants $c_2 > c_1 > 0$, $\gamma_2 \ge \gamma_1 > 0$ or $\gamma_2 = \gamma_1 = 0$, and $d \ge d_0 > 0$ such that

$$c_1\Big(\frac{R}{r}\Big)^{d_0} \le \frac{V(R)}{V(r)} \le c_2\Big(\frac{R}{r}\Big)^d \qquad \text{for every } \ 0 < r < R < \infty. \tag{1.11}$$

and

$$\phi(r) = \phi_1(r)\psi(r) \qquad \text{for } r > 0, \tag{1.12}$$

where $\psi$ is an increasing function on $[0, \infty)$ with $\psi(r) = 1$ for $0 < r \le 1$ and

$$c_1 e^{\gamma_1 r} \le \psi(r) \le c_2 e^{\gamma_2 r} \qquad \text{for every } \ 1 < r < \infty,$$

and $\phi_1$ is a strictly increasing function on $[0, \infty)$ with $\phi_1(0) = 0$, $\phi_1(1) = 1$ and satisfies the following: there exist constants $c_2 > c_1 > 0$, $c_3 > 0$, and $\beta_2 \ge \beta_1 > 0$ such that

$$c_1\Big(\frac{R}{r}\Big)^{\beta_1} \le \frac{\phi_1(R)}{\phi_1(r)} \le c_2\Big(\frac{R}{r}\Big)^{\beta_2} \qquad \text{for every } 0 < r < R < \infty, \tag{1.13}$$

$$\int_0^r \frac{s}{\phi_1(s)}\, ds \le c_3 \frac{r^2}{\phi_1(r)} \qquad \text{for every } r > 0. \tag{1.14}$$

**Remark 1.1** Note that conditions (1.11) and (1.13) are equivalent to the existence of constants $c_4, c_5 > 1$ and $L_0 > 1$ such that for every $r > 0$,

$$c_4\phi_1(r) \le \phi_1(L_0 r) \le c_5\,\phi_1(r) \quad \text{and} \quad c_4 V(r) \le V(L_0 r) \le c_5\, V(r).$$

Assume further that there is a metric space $X \supset F$, and $\rho(\cdot,\cdot)$ can be extended to be a metric on $X$ with dilation for $F$, i.e. there are constants $c_6 \ge 1$ such that for every $x, y \in F$ and $\delta > 0$, $\delta^{-1}x, \delta^{-1}y \in X$ with

$$c_6^{-1}\delta^{-1}\rho(x,y) \le \rho(\delta^{-1}x, \delta^{-1}y) \le c_6\, \delta^{-1}\rho(x,y). \tag{1.15}$$

Clearly the above condition is satisfied if $F \subset \mathbb{R}^n$ as we can take $X$ to be $\mathbb{R}^n$. See [16] for a non-Euclidean example of $F$ satisfying the above condition.

The main result of this paper is the following heat kernel estimates. The inverse function of the strictly increasing function $t \mapsto \phi(t)$ is denoted by $\phi^{-1}(t)$.

**Theorem 1.2** *Assume that $\mu(F) = \infty$ and that the conditions (1.1), (1.8)-(1.9), (1.11)-(1.12) and (1.15) hold. Then there is a conservative Feller process $Y$ associated with $(\mathcal{E}, \mathcal{F})$ that starts from every point in $F$. Moreover the process $Y$ has a continuous transition density function on $(0,\infty) \times F \times F$ with respect to the measure $\mu$, which has the following estimates. There are positive constants $c_1 > 0$, $c_2 > 0$ and $C \ge 1$ such that*

$$C^{-1}\left(\frac{1}{V(\phi^{-1}(t))} \wedge \frac{t}{V(\rho(x,y))\phi(c_1\rho(x,y))}\right) \le p(t,x,y) \le C\left(\frac{1}{V(\phi^{-1}(t))} \wedge \frac{t}{V(\rho(x,y))\phi(c_2\rho(x,y))}\right),$$

*for every $t \in (0,1]$ and $x, y \in F$. Moreover, when $\gamma_1 = \gamma_2 = 0$ in (1.12), then the above heat kernel estimates hold for every $t > 0$ and $x, y \in F$.*

**Remark 1.3** (i) Conservativeness of $Y$ can be established under weaker assumptions (1.1)-(1.3) and (1.7)-(1.9). In particular, we do not need assumptions (1.15) nor $\mu(F) = \infty$ for conservativeness. See Theorem 4.7 below.

(ii) Concrete examples are given in Example 2.3 for which conditions (1.8)-(1.9), and (1.11)-(1.12) are satisfied. In particular, they include mixtures of stable processes.

(iii) The existence of a continuous heat kernel for $Y$ in Theorem 1.2 implies that we can refine the process $Y$ so that it starts from *every point* in $F$ (cf. [14]).

(iv) We point out that the heat estimate given Theorem 1.2 is only true for small time $t$ in the case of $\gamma_2 \ge \gamma_1 > 0$. This is illustrated by the example of relativistic 1-stable process on $\mathbb{R}^d$ given in Example 2.4 below, where the heat kernel behavior when $t \to \infty$ is quite different. We will investigate the large time or global time heat kernel behavior in a separate paper.

The remainder of the paper is organized as follows. Section 2 presents several inequalities that will be used in later sections as well as some examples. Nash's inequality and on-diagonal upper bound estimate is established in Section 3. The main results of this paper is proved in Section 4. To establish the off-diagonal heat kernel estimate for $p(t,x,y)$ when $\rho(x,y)\geq 1$, we decompose the Dirichlet form of $Y$ into two parts: forms with jumping intensity kernels $J(\xi,\eta)\mathbf{1}_{\{\rho(\xi,\eta)\leq\lambda\}}$ and $J(\xi,\eta)\mathbf{1}_{\{\rho(\xi,\eta)>\lambda\}}$, where $\lambda$ is carefully chosen, proportional to the distance between $x$ and $y$. The heat kernel $p^*(t,x,y)$ for the process $Y^*$ corresponding to the jumping intensity kernel $J(\xi,\eta)\mathbf{1}_{\{\rho(\xi,\eta)\leq\lambda\}}$ can be estimated from above by using a method from Carlen-Kusuoka-Stroock [9]. Then the upper bound estimate for $p(t,x,y)$ can be obtained from that of $p^*(t,x,y)$ due to the Meyer's construction of $Y$ from $Y^0$ and vice versa. To establish the off-diagonal heat kernel estimate for $p(t,x,y)$ when $\rho(x,y)<1$ in the case of $\gamma_1=\gamma_2=0$, we use a scaling argument. The heat kernel estimate for $p(t,x,y)$ in the case of $\gamma_2\geq\gamma_1>0$ is then derived from the upper bound estimate of $p(t,x,y)$ in the case of $\gamma_1=\gamma_2=0$. We remark that as oppose to the approaches in [6] and [11], the upper bound heat kernel estimate is obtained in this paper without having to establish the parabolic Harnack inequality first. From the upper bound heat kernel estimate, we can then derive certain hitting probability estimates (or tightness results) and prove parabolic Harnack inequality. These results yield heat kernel lower bound estimate. The parabolic Harnack inequality implies the generalized Hölder continuity of $p(t,x,y)$.

Throughout this paper, we will use $c$, with or without subscripts, to denote strictly positive finite constants whose values are insignificant and may change from line to line. For $r\in[1,\infty]$, we will use $\|f\|_r$ to denote the $L^r$-norm in $L^r(F,\mu)$.

## 2 Preliminaries

The following Lemma will be needed later.

**Lemma 2.1** *Assume that conditions (VD) and (1.2)-(1.4) hold. There exist positive constants $c_1$ and $c_2$ such that*

*(i)* $M_0(r):=\sup_\eta\int_{B(\eta,r)^c}J(\eta,\xi)\mu(d\xi)<c_1/\phi(r)$ *for all* $r>0$.

*(ii)* $M_1(r):=\sup_\eta\int_{B(\eta,r)}\rho(\eta,\xi)^2J(\eta,\xi)\mu(d\xi)<c_2\int_0^r\frac{s}{\phi(s)}ds$ *for all* $r>0$.

**Proof.** For simplicity, define $V(x,r):=\mu(B(x,r))$.

(i) By (1.4) and (VD), we have

$$
\begin{aligned}
\int_{B(\eta,r)^c} J(\eta,\xi)\mu(d\xi) &\le c_1 \int_{B(\eta,r)^c} \frac{1}{\phi(\rho(\eta,\xi))\, V(\eta,\rho(\eta,\xi))}\mu(d\xi) \\
&\le c_2 \sum_{i=0}^{\infty} \frac{1}{\phi(2^i r)\, V(\eta,2^i r)} \left( V(\eta, 2^{i+1}r) - V(\eta, 2^i r)\right) \\
&\le c_2 \sum_{i=0}^{\infty} \frac{1}{\phi(2^i r)\, V(\eta,2^i r)} \cdot (c-1)V(\eta,2^i r) \\
&\le \frac{c_3}{\phi(r)} \sum_i \frac{\phi(r)}{\phi(2^i r)} \\
&\le \frac{c_4}{\phi(r)} \sum_i 2^{-i\beta_1} \\
&\le \frac{c_5}{\phi(r)},
\end{aligned}
$$

where the lower bound in (1.2) is used in the second to the last inequality.

(ii) can be established similarly.

$$
\begin{aligned}
\int_{B(\eta,r)} \rho(\eta,\xi)^2 J(\eta,\xi)\mu(d\xi) &\le c_1 \int_{B(\eta,r)} \rho(\eta,\xi)^2 \frac{1}{\phi(\rho(\eta,\xi))\, V(\eta,\rho(\eta,\xi))}\mu(d\xi) \\
&\le c_2 \sum_{i=0}^{\infty} (2^{-i}r)^2 \frac{V(\eta, r2^{-i}) - V(\eta, 2^{-i-1}r)}{\phi(2^{-(i+1)}r)\, V(\eta, 2^{-i-1}r)} \\
&\le c_2 \sum_{i=0}^{\infty} (2^{-i}r)^2 \frac{(c-1)V(\eta, 2^{-i-1}r)}{\phi(2^{-i-1}r)\, V(\eta, 2^{-i-1}r)} \\
&\le c_3 \sum_{i=0}^{\infty} \frac{(2^{-i}r)^2}{\phi(2^{-i-1}r)} \\
&\le c_4 \int_0^r \frac{s}{\phi(s)}\, ds.
\end{aligned}
$$

□

Recall the definition of $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ given by (1.5)-(1.6).

**Proposition 2.2** *Under the assumptions (VD), (1.2), (1.3) and (1.4), $\mathcal{D}(\mathcal{E})$ is dense in $C_c(F)$. Moreover, if we define $\mathcal{F} := \overline{\mathcal{D}(\mathcal{E})}^{\mathcal{E}_1}$, then $(\mathcal{E}, \mathcal{F})$ is a regular Dirichlet form on $L^2(F,\mu)$.*

**Proof.** For each $x \neq y \in F$, let $r := \rho(x,y)$ and define

$$\psi(\xi) = 1 - \frac{\rho(\xi, x) \wedge r}{r}, \quad \xi \in F.$$

Clearly $\psi \in C_c(F)$, $\mathrm{supp}[\psi] \subset B(x,r)$ and $|\psi(\xi) - \psi(\eta)| \le \rho(\eta,\xi)/r$. So by Lemma 2.1,

$$\begin{aligned}
\mathcal{E}(\psi) &= \int_{B(x,r)} \int_{B(x,r)} (\psi(\xi)-\psi(\eta))^2 J(\xi,\eta)\mu(d\xi)\mu(d\eta) \\
&\quad +2\int_{B(x,r)^c} \mu(d\xi) \int_{B(x,r)} \psi(\eta)^2 J(\xi,\eta)\mu(d\eta) \\
&\le \frac{1}{r^2}\int_{B(x,r)}\int_{B(x,r)} \rho(\xi,\eta)^2 J(\xi,\eta)\mu(d\xi)\mu(d\eta) + 2M_0(r)\int_{B(x,r)} \psi(\eta)^2\mu(d\eta) \\
&\le \frac{M_1(r)}{r^2}\mu(B(x,r)) + 2M_0(r)\mu(B(x,r)) \\
&< \infty.
\end{aligned}$$

Thus $\psi \in \mathcal{D}(\mathcal{E})$. Since this holds for all $x \neq y \in F$, using Stone-Weierstrass' theorem we see that $\mathcal{D}(\mathcal{E})$ is dense in $C_c(F)$. On the other hand, since $F$ is a locally compact separable metric space, $C_c(F)$ is dense in $L^2(F,\mu)$. Thus, using Example 1.2.4 in [14] for instance, we can show that $(\mathcal{E},\mathcal{F})$ is a regular Dirichlet form on $L^2(F,\mu)$. □

**Examples 2.3** We now give some examples such that condition (1.12) is satisfied.

(i) If there is $0 < \alpha_1 < \alpha_2 < 2$ and a probability measure $\nu$ on $[\alpha_1,\alpha_2]$ such that

$$\phi(r) := \int_{\alpha_1}^{\alpha_2} r^\alpha \, \nu(d\alpha),$$

then condition (1.12) is satisfied with $\gamma_1 = \gamma_2 = 0$. Clearly, $\phi$ is a continuous strictly increasing function with $\phi(0) = 0$ and $\phi(1) = 1$. The condition (1.12) is satisfied with $\gamma_1 = \gamma_2 = 0$ because

$$\frac{1}{2^{\alpha_1}} \le \frac{\phi(r)}{\phi(2r)} \le \frac{1}{2^{\alpha_2}} \qquad \text{for any } r > 0.$$

Note that, in this case,

$$J(x,y) \asymp \frac{1}{V(\rho(x,y)) \int_{\alpha_1}^{\alpha_2} \rho(x,y)^\alpha \nu(d\alpha)}.$$

(ii) Similarly, condition (1.12) is satisfied with $\gamma_1 = \gamma_2 = 0$ if

$$\phi(r) := \left( \int_{\alpha_1}^{\alpha_2} r^{-\alpha} \, \nu(d\alpha) \right)^{-1},$$

where $\nu$ is a probability measure on $[\alpha_1,\alpha_2] \subset (0,2)$. In this case,

$$J(x,y) \asymp \int_{\alpha_1}^{\alpha_2} \frac{1}{V(\rho(x,y)\rho(x,y)^\alpha} \nu(d\alpha).$$

A particular case is when $\nu$ is a discrete measure. When $F=\mathbb{R}^n$, $\rho(x,y)$ is the Euclidean distance between $x$ and $y$, and $\mu$ is the Lebesgue measure on $\mathbb{R}^d$, Theorem 1.2 in particular gives the heat kernel estimate for Lévy processes on $\mathbb{R}^d$ which are linear combinations of independent of independent symmetric $\alpha$-stable processes. Of course, our theorem holds much more generally, even in the case of $F=\mathbb{R}^n$.

(iii) $\phi(r)=r^{\alpha(r)}$, where $0\le \alpha(r)\le \beta<2$, $a(r)\log r$ is increasing and $(a(2r)-a(r))\log r$ is bounded. $a(r)=c-\frac{c}{|r-1|+1}$ is a such an example for $c\in(0,2)$. This case corresponds to the jumping density

$$J(x,y)\asymp \frac{1}{V(\rho(x,y)\rho(x,y)^{\alpha(\rho(x,y))}}.$$

(iv) In the above examples, condition (1.12) is still satisfied if one adds a multiplicative term $e^{c_0(r)r}$, where $c_0(r)$ is a symmetric function bounded between two positive constant $\gamma_1$ and $\gamma_2$. □

**Examples 2.4** Let $Y=\{Y_t,t\ge 0\}$ be the relativistic $\alpha$-stable processes on $\mathbb{R}^d$ with mass $m>0$. That is, $\{Y_t,t\ge 0\}$ is a Lévy process on $\mathbb{R}^d$ with

$$\mathbb{E}^0[\exp(i\langle\xi,Y_t\rangle)]=\exp\left(t\left(m^\alpha-(|\xi|^2+m^2)^{\alpha/2}\right)\right),$$

where $m>0$, $\alpha\in(0,2)$. It is shown in [12] that the corresponding jumping intensity satisfies

$$J(x,y)\asymp\frac{\Psi(m|x-y|)}{|x-y|^{d+\alpha}},$$

where $\Psi(r)\asymp e^{-r}(1+r^{(d+\alpha-1)/2})$ near $r=\infty$, and $\Psi(r)=1+\Psi''(0)r^2/2+o(r^4)$ near $r=0$. So condition (1.12) is satisfied with $\gamma_1>0$ for the jumping intensity kernel for every relativistic $\alpha$-stable processes on $\mathbb{R}^d$.

When $\alpha=1$, the process is called a relativistic Hamiltonian process. In this case, the heat kernel can be written as

$$p(t,x,y)=\frac{t}{(2\pi)^d\sqrt{|x-y|^2+t^2}}\int_{\mathbb{R}^d} e^{mt}e^{-\sqrt{(|x-y|^2+t^2)(|z|^2+m^2)}}dz,$$

see [15], also [19]. For simplicity, take $m=1$. It can be shown (to appear somewhere else) that for every $t>0$ and $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$,

$$\begin{aligned}&\frac{c_1t}{(|x-y|+t)^{d+1}}\left(1\vee(|x-y|+t)^{d/2}\right)e^{-c_2\frac{|x-y|^2}{\sqrt{|x-y|^2+t^2}}}\\ \le\ & p(t,x,y)\\ \le\ &\frac{c_3t}{(|x-y|+t)^{d+1}}\left(1\vee(|x-y|+t)^{d/2}\right)e^{-c_4\frac{|x-y|^2}{\sqrt{|x-y|^2+t^2}}}.\end{aligned}$$

This in particular implies that for every fixed $t_0 > 0$, there exist $c_1, \cdots, c_4 > 0$ which depend on $t_0$ such that

$$c_1 \left( t^{-d} \wedge \frac{t}{|x-y|^{d+1}} \right) e^{-c_2|x-y|} \le p(t,x,y) \le c_3 \left( t^{-d} \wedge \frac{t}{|x-y|^{d+1}} \right) e^{-c_4|x-y|}$$

for every $t \in (0, t_0]$ and $x, y \in \mathbb{R}^d$. □

As mentioned previously, the above example shows that the estimate given Theorem 1.2 is only true for small time $t$ in the case of $\gamma_2 \ge \gamma_1 > 0$. The heat kernel behavior when $t \to \infty$ is different.

# 3 Nash's inequality and on-diagonal heat kernel upper bound estimate

Throughout this section, we will assume conditions (1.2)-(1.3), and (1.7)-(1.9). Recall the definition of the regular Dirichlet form $(\mathcal{E}, \mathcal{F})$ on $L^2(F, \mu)$ given by (1.5)-(1.6).

**Theorem 3.1** *There are positive constants $c_1, c_2 > 0$, depending only on the multiplicative constants in (1.7)-(1.9), such that for every $u \in \mathcal{F}$ with $\|u\|_1 = 1$, we have*

$$\theta(\|u\|_2^2) \le c_2 \, \mathcal{E}(u,u), \tag{3.1}$$

*where $\theta(r) := \frac{r}{\phi(V^{-1}(c_1 r^{-1}))}$ and $V^{-1}$ is the inverse function of $r \mapsto V(r)$.*

**Proof.** For $r > 0$, define

$$u_r(x) := \frac{1}{\mu(B(x,r))} \int_{B(x,r)} u(z)\mu(dz) \qquad \text{for } x \in F.$$

We have

$$\|u_r\|_\infty \le c_1 \|u\|_1 / V(r) \quad \text{ and } \quad \|u_r\|_1 \le c_1 \|u\|_1.$$

Thus

$$\|u_r\|_2^2 \le \|u_r\|_\infty \|u_r\|_1 \le c_2 \|u\|_1^2 / V(r). \tag{3.2}$$

So for $u \in \mathcal{F}$ with $\|u\|_1 = 1$,

$$\begin{aligned}
\|u\|_2^2 &\leq 2\|u - u_r\|_2^2 + 2\|u_r\|_2^2 \\
&\leq 2\int_F \left( \frac{1}{\mu(B(x,r))} \int_{B(x,r)} (u(x) - u(y))^2 \mu(dy) \right) \mu(dx) + \frac{2c_2\|u\|_1^2}{V(r)} \\
&\leq c_3 \int_F \left( \frac{1}{\mu(B(x,r))} \int_{B(x,r)} (u(x) - u(y))^2 J(x,y)\, \phi(\rho(x,y)) V(\rho(x,y))\, \mu(dy) \right) \mu(dx) + \frac{2c_2}{V(r)} \\
&\leq \frac{c_4}{V(r)} \phi(r) V(r) \int_F \left( \int_{B(x,r)} (u(x) - u(y))^2 J(x,y) \mu(dy) \right) \mu(dx) + \frac{2c_2}{V(r)} \\
&\leq c_5 \left( \phi(r)\mathcal{E}(u,u) + \frac{1}{V(r)} \right). \qquad (3.3)
\end{aligned}$$

To minimize the right hand side, let $\phi(r_0)\mathcal{E}(u,u) = 1/V(r_0)$; that is,

$$(\phi(r_0)V(r_0))^{-1} = \mathcal{E}(u,u). \qquad (3.4)$$

Note that it follows from (1.2) and (1.7), $r \to (\phi(r)V(r))^{-1}$ is a strictly decreasing continuous function with

$$\lim_{r\to 0+} (\phi(r)V(r))^{-1} = +\infty \qquad \text{and} \qquad \lim_{r\to\infty} (\phi(r)V(r))^{-1} = 0$$

and so (3.4) has a unique solution $r_0$. So from (3.3), we see that $\|u\|_2^2 \leq 2c_5/V(r_0)$, or equivalently,

$$r_0 \leq V^{-1}(2c_5 \|u\|_2^{-2}).$$

Since $\phi$ is a continuous increasing function,

$$V(r_0)\phi(r_0) \leq 2c_5 \|u\|_2^{-2} \phi(V^{-1}(2c_5 \|u\|_2^{-2})) = \frac{c_6}{\theta(\|u\|_2^2)}.$$

So by (3.4)

$$\theta(\|u\|_2^2) \leq c_6 \mathcal{E}(u,u).$$

This proves the theorem. □

We know by Proposition 2.2 that $(\mathcal{E},\mathcal{F})$ is a regular Dirichlet form on $L^2(F,\mu)$. So there exists $\mathcal{N}_0 \subset F$ having zero capacity with respect to the Dirichlet form $(\mathcal{E},\mathcal{F})$ and there is a Hunt process $(Y,\mathbb{P}^x)$ with state space $F \setminus \mathcal{N}_0$ such that for every $f \in L^2(F,\mu)$ and $t > 0$, $x \mapsto \mathbb{E}^x[f(X_t)]$ is a quasi-continuous version of $T_t f$, where $\{T_t, t \geq 0\}$ is the semigroup associated with the closed form $(\mathcal{E},\mathcal{F})$. In fact, the exceptional set $\mathcal{N}_0$ can be chosen so that the Hunt process $Y$ can start from any point in $F \setminus \mathcal{N}_0$ and that once it starts from $F \setminus \mathcal{N}_0$ the process $Y$ together with its left hand limits takes values in $F \setminus \mathcal{N}_0$ up to and strictly before its lifetime $\zeta$. Such a set $\mathcal{N}_0$ is called *a properly exceptional set* of $Y$ (or, equivalently, of $(\mathcal{E}, \mathcal{F})$) and it always has zero $\mu$-measure. For simplicity, sometimes we just say that $Y$ is a Hunt process associated with $(\mathcal{E}, \mathcal{F})$ starting from quasi-everywhere in $F$. For more on terminologies and properties of Dirichlet forms, we refer the reader to [14].

Let $P(t, x, dy)$ be the transition probability for the Hunt process $Y$ associated with $(\mathcal{E}, \mathcal{F})$. It is well-known that the Nash's type inequality (3.1) yields that for $\mu$-a.e. $x \in F$, $P(t, x, dy)$ has a density function $p(t, x, y)$ with respect to measure $\mu$ and that an almost everywhere on-diagonal heat kernel upper bound estimate holds for $p(t, x, y)$. Using the argument for the proof of Theorem 1.2 in [1] (see also [11]), we can show that in fact a quasi-continuous kernel $p(t, x, y)$ exists for $P(t, x, dy)$ and that the estimate holds quasi-everywhere.

**Theorem 3.2** *There is a properly exceptional set $\mathcal{N} \supset \mathcal{N}_0$ of $Y$, a positive symmetric kernel $p(t, x, y)$ defined on $(0, \infty) \times (F \setminus \mathcal{N}) \times (F \setminus \mathcal{N})$, and positive constants $c_1$ and $c_2$, depending on the multiplicative constants in (1.2)-(1.3) and(1.7)-(1.9), such that $P(t, x, dy) = p(t, x, y)\mu(dy)$, and*

$$p(t, x, y) \le \frac{c_2}{V(\phi^{-1}(c_1 t))} \qquad \textit{for every } x, y \in F \setminus \mathcal{N} \textit{ and for every } t > 0. \tag{3.5}$$

*Moreover, for every $t > 0$ and $y \in F \setminus \mathcal{N}$, $x \mapsto p(t, x, y)$ is quasi-continuous on $F$.*

**Proof.** For $f \ge 0$, define $P_t f(x) := \mathbb{E}_x\left[f(Y_t)\right]$ for $x \in F \setminus \mathcal{N}_0$. Note that under the condition (1.2) and (1.7)-(1.8), the function $r \mapsto 1/\theta(r)$ is integrable at $r = \infty$, where $\theta$ is the function in Theorem 3.1. Thus according to [13, Proposition II.1], the Nash-type inequality (3.1) implies that

$$\|P_t f\|_\infty \le m(t)\, \|f\|_1 \qquad \text{for every } t > 0 \text{ and } f \in L^1(F, \mu),$$

where $m(t)$ is the inverse function of

$$h(t) := \int_t^\infty \frac{c}{\theta(x)} dx.$$

Using the explicit expression of $\theta$ in Theorem 3.1,

$$h(t) = \int_t^\infty \frac{c\phi(V(c_1 x^{-1}))}{x} dx = c_2 \int_0^{c_1 t^{-1}} \frac{\phi(V^{-1}(y))}{y} dy.$$

Observe that

$$\int_0^{c_1 t^{-1}} \frac{\phi(V^{-1}(y))}{y} dy \le \sum_{k=0}^\infty \frac{\phi(V^{-1}(2^{-k} c_1 t^{-1}))}{2^{-k-1} c_1 t^{-1}} \left(2^{-k} c_1 t^{-1} - 2^{-k-1} c_1 t^{-1}\right) = \sum_{k=0}^\infty \phi(V^{-1}(2^{-k} c_1 t^{-1})).$$

By (1.7), there exists $c > 1$ such that $cV^{-1}(s) \le V^{-1}(2s)$ for all $s > 0$. Thus,

$$\sum_{k=0}^\infty \phi(V^{-1}(2^{-k} c_1 t^{-1})) \le \sum_{k=0}^\infty \phi(c^{-k} V^{-1}(c_1 t^{-1})) \le c_2 \phi(V^{-1}(c_1 t^{-1})) \sum_{k=0}^\infty c^{-k\beta_1} \le c_3 \phi(V^{-1}(c_1 t^{-1})),$$

where (1.2) is used in the second inequality.

Thus $h(t) \le c_4 \phi(V^{-1}(c_1 t^{-1}))$ for every $t > 0$. Taking the inverse function, we obtain $m(t) \le c_1/V(\phi^{-1}(t/c_4))$ and so

$$\|P_t f\|_\infty \le \frac{c_1}{V(\phi^{-1}(t/c_4))} \|f\|_1 \qquad \text{for every } t > 0 \text{ and } f \in L^1(F,\mu).$$

Now the conclusion of the theorem follows from a straightforward modification of the proof of Theorem 1.2 in [1]. □

For simplicity, we call a symmetric kernel $p(t,x,y)$ having the properties in Theorem 3.2 except the estimate (3.5) *quasi-continuous*.

# 4 Heat kernel estimates

In the first two subsections, we will prepare several estimates that will be used in the proof of Theorem 1.2. We then prove the upper and lower bounds in §4.3–4.9 for $x, y \in F \setminus \mathcal{N}$ for the properly exceptional set $\mathcal{N}$. In §4.10, we use the Hölder continuity and obtain the heat kernel estimates for all $x, y \in F$.

We will need to decompose the Dirichlet form $(\mathcal{E}, \mathcal{F})$ into two forms, a form with small jumps and the other with large jumps. Then we apply the Meyer's construction ([20]) as in [1, 2].

## 4.1 Meyer's construction

The following construction of Meyer [20] for jump processes will be used in our approach. Suppose we have jump intensity kernels $J(x,y), J_0(x,y)$ on $F \times F$ such that their corresponding pure jump Dirichlet forms given in terms of (1.5) with $\mathcal{F} = \overline{\mathcal{D}(\mathcal{E})}^{\mathcal{E}_1}$ are regular on $F$. Let $Y = \{Y_t, t \ge 0, \mathbb{P}^x, x \in F \setminus \mathcal{N}\}$ and $Y^{(0)} = \{Y_t^{(0)}, t \ge 0, \mathbb{P}^x, x \in F \setminus \mathcal{N}_0\}$ be the processes corresponding to the Dirichlet forms whose Lévy densities are $J(x,y), J_0(x,y)$. Here $\mathcal{N}$ and $\mathcal{N}_0$ are the properly exceptional sets of $Y$ and $Y^{(0)}$, respectively. Suppose that $J_0(x,y) \le J(x,y)$ and

$$\mathcal{J}(x) := \int_F (J(x,y) - J_0(x,y))\mu(dy) \le c,$$

for all $x \in F$. Let

$$J_1(x,y) := J(x,y) - J_0(x,y) \qquad \text{and} \qquad q(x,y) = \frac{J_1(x,y)}{\mathcal{J}(x)}. \tag{4.1}$$

Then we can construct a process $Y$ corresponding to the jump kernel $J$ from $Y^{(0)}$ as follows. Let $S_1$ be an exponential random variable of parameter 1 independent of $Y^{(0)}$, let $C_t = \int_0^t \mathcal{J}(Y_s^{(0)})\, ds$, and let $U_1$ be the first time that $C_t$ exceeds $S_1$. We let $Y_s = Y_s^{(0)}$ for $0 \le s < U_1$.

At time $U_1$ we introduce a jump from $Y_{U_1-}$ to $Z_1$, where $Z_1$ is chosen at random according to the distribution $q(Y_{U_1-}, y)$. We set $Y_{U_1} = Z_1$, and repeat, using an independent exponential $S_2$, etc. Since $\mathcal{J}(x)$ is bounded, only finitely many new jumps are introduced in any bounded time interval. In [20] it is proved that the resulting process corresponds to the kernel $J$. See also [17]. Note that if $\mathcal{N}_0$ is the properly exceptional set corresponding to $Y^{(0)}$, then this construction gives that the properly exceptional set $\mathcal{N}$ for $Y$ can be chosen to be a subset of $\mathcal{N}_0$.

Conversely, we can also remove a finite number of jumps from a process $Y$ to obtain a new process $Y^{(0)}$. For simplicity, assume that $J_0(x,y)J_1(x,y) = 0$. Suppose one starts with the process $Y$ (associated with $J$), runs it until the stopping time $S_1 = \inf\{t : J_1(Y_{t-}, Y_t) > 0\}$, and at that time restarts $Y$ at the point $Y_{S_1-}$. Suppose one then repeats this procedure over and over. Meyer [20] proves that the resulting process $Y^{(0)}$ will correspond to the jump kernel $J_0$. In this case $\mathcal{N}_0 \subset \mathcal{N}$.

Assume that the processes $Y$ and $Y^{(0)}$ have transition density functions $p(t,x,y)$ and $p^{(0)}(t,x,y)$, respectively. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be the filtration generated by the process $Y^{(0)}$. The following lemma is shown in [1, Lemma 2.4] and in [2, Lemma 3.2].

**Lemma 4.1** (i) *For any $A \in \mathcal{F}_t$,*

$$\mathbb{P}^x \left( \{Y_s = Y_s^{(0)} \textit{ for all } 0 \leq s \leq t\} \cap A \right) \geq e^{-t\|\mathcal{J}\|_\infty}\, \mathbb{P}^x(A).$$

(ii) *If $\|J_1\|_\infty < \infty$, then*

$$p(t,x,y) \leq p^{(0)}(t,x,y) + t\|J_1\|_\infty.$$

## 4.2 Scaling

We now return to the Dirichlet form $(\mathcal{E}, \mathcal{F})$ defined by (1.5), where the jumping intensity kernel is specified by (1.4). Throughout the remaining of Section 4, with the exception of §4.6, we assume that $\mu(F) = \infty$ and the conditions (1.1), (1.8)-(1.9), (1.11)-(1.12) and (1.15) hold. §4.6 deals with the conservativeness of the jump process $Y$, which can be proved under weaker assumptions (1.1)-(1.3) and (1.7)-(1.9). In particular, we do not need assumptions (1.15) nor $\mu(F) = \infty$ in §4.6.

To derive a sharp upper bound heat kernel estimate for $(\mathcal{E}, \mathcal{F})$, we will use a scaling technique. This is where condition (1.15) is used.

For any $\delta > 0$ , let $G^{(\delta)} := \delta^{-1}F = \{\delta^{-1}x : x \in F\}$, and let $\nu^{(\delta)}$ be the measure on $G^{(\delta)}$ defined by

$$\nu^{(\delta)}(A) = \frac{\mu(\delta A)}{V(\delta)} \qquad \text{for every measurable } A \subset G^{(\delta)}.$$

When there is no confusion possible, the ball in $G^{(\delta)}$ centered at $x$ with radius $r$ will still be denoted by $B(x,r)$. In this subsection, we will deal with processes and Dirichlet forms on $G^{(\delta)}$. For notational simplification, we will suppress $\delta$ from $G^{(\delta)}$, $\nu^{(\delta)}$, $\cdots$, etc, and denote them by $G$, $\nu$, $\cdots$.

To derive an upper bound estimate for the transition density function $p(t,x,y)$ for the Hunt process $Y$ on $F$, for suitable $\delta>0$, we look at the scaled process

$$Y_t^{(\delta)} := \delta^{-1} Y_{\phi(\delta)t}, \qquad t \geq 0.$$

Clearly $Y^{(\delta)}$ is a $\nu$-symmetric Hunt process on $G$ and it starts from every point in $G \setminus \delta^{-1}\mathcal{N}$, where $\mathcal{N}$ is the properly exceptional set of $Y$ on $F$ in Theorem 3.2. Since

$$\int_A h(\xi)\nu(d\xi) = \frac{1}{V(\delta)} \int_{\delta A} h(\delta^{-1}x)\mu(dx) \tag{4.2}$$

for every measurable $A \subset G$ and every $\nu$-integrable function $h$ on $A$, it is easy to check the following (cf. [11]).

**Lemma 4.2** *The Dirichlet form $(W, \mathcal{D}(W))$ of $Y^{(\delta)}$ in $L^2(G,\nu)$ is given by*

$$\begin{aligned} W(u,u) &:= \int_{G\times G} (u(\xi)-u(\eta))^2 J^{(\delta)}(\xi,\eta)\nu(d\xi)\nu(d\eta), \\ D(W) &= \overline{\{u \in C_c(G) \text{ with } W(u,u)<\infty\}}^{W_1}, \end{aligned}$$

*where*

$$J^{(\delta)}(\xi,\eta) = \phi(\delta)V(\delta)J(\delta\xi,\delta\eta) \qquad \text{for every } \xi,\eta \in G,$$

*and $W_1(u,u) := W(u,u) + \int_G u(x)^2\nu(dx)$. The Dirichlet form $(W, \mathcal{D}(W))$ is regular in $L^2(G,\nu)$.*

In view of (1.9) and (1.15), there are constants $c \geq 1$ independent of $\delta > 0$ such that

$$c^{-1}\frac{1}{V^{(\delta)}(\rho(\xi,\eta))\phi^{(\delta)}(\rho(\xi,\eta))} \leq J^{(\delta)}(\xi,\eta) \leq c\,\frac{1}{V^{(\delta)}(\rho(\xi,\eta))\phi^{(\delta)}(\rho(\xi,\eta))} \qquad \text{for } \xi,\eta \in G,$$

where

$$V^{(\delta)}(r) := V(\delta r)/V(\delta) \quad \text{ and } \phi^{(\delta)}(r) := \phi(\delta r)/\phi(\delta). \tag{4.3}$$

We note that $\phi^{(\delta)}$ and $V^{(\delta)}$ satisfy the same condition (1.2), (1.3) and (1.7) by $\phi$ and $V$, respectively, with the same multiplicative constants. Thus results proved in §2 and §3 hold for $(W, \mathcal{D}(W))$ and $Y^{(\delta)}$ with the multiplicative constants independent of $\delta > 0$.

Note that the transition density function $q^{(\delta)}(t,x,y)$ of $Y^{(\delta)}$ with respect to the measure $\nu$ is related to that of $Y$ by the formula

$$q^{(\delta)}(t,\xi,\eta) = V(\delta)\, p(\phi(\delta)t, \delta\,\xi, \delta\,\eta) \qquad \text{for } t>0 \text{ and } \xi,\eta \in G\setminus\delta^{-1}\mathcal{N}. \tag{4.4}$$

### 4.3 Decomposition of quadratic forms

To derive heat kernel estimate, we need to decompose the Dirichlet form $(W^{(\lambda)}, \mathcal{D}(W))$ into two quadratic forms, one with small jumps and the other with large jumps, and then use Meyer's construction.

Let $\lambda > 0$ and consider the bilinear form $(W^{(\delta,\lambda)}, \mathcal{D}(W))$ on $L^2(G, \nu)$ defined by

$$W^{(\delta,\lambda)}(v, v) = \int_G \int_G (v(\xi) - v(\eta))^2 \, \mathbf{1}_{\{\rho(\eta,\xi)\leq\lambda\}} J^{(\delta)}(\xi, \eta) \, \nu(d\eta)\nu(d\xi). \tag{4.5}$$

We denote

$$J^{(\delta,\lambda)}(\xi, \eta) := \mathbf{1}_{\{\rho(\xi,\eta)\leq\lambda\}} \, J^{(\delta)}(\xi, \eta) \quad \text{and} \quad J^{(\delta)}_\lambda(\xi, \eta) := \mathbf{1}_{\{\rho(\xi,\eta)>\lambda\}} \, J^{(\delta)}(\xi, \eta).$$

By Lemma 2.1(i), there is a constant $c > 0$, independent of $\delta > 0$ and $\lambda > 0$, such that

$$\int_G J^{(\delta)}_\lambda(\xi, \eta)\nu(d\eta) \leq \frac{c}{\phi^{(\delta)}(\lambda)} \qquad \text{for every } \xi \in G.$$

Therefore we have for $v \in \mathcal{D}(W)$,

$$0 \leq W(v, v) - W^{(\delta,\lambda)}(v, v) \leq 4 \int_G v(\xi)^2 \left( \int_G J^{(\delta)}_\lambda(\xi, \eta)\nu(d\eta) \right) \nu(d\xi) \leq \frac{c}{\phi^{(\delta)}(\lambda)} \int_G v(\xi)^2 \nu(d\xi),$$

where $c > 0$ is a constant independent of $\delta > 0$ and $\lambda > 0$. Thus

$$\left( \frac{c}{\phi^{(\delta)}(\lambda)} + 1 \right)^{-1} W_1(v, v) \leq W_1^{(\delta,\lambda)}(v, v) \leq W_1(v, v) \qquad \text{for every } v \in \mathcal{D}(W). \tag{4.6}$$

It follows then $(W^{(\delta,\lambda)}, \mathcal{D}(W))$ is a regular Dirichlet form on $L^2(G, \nu)$ and so there is a Hunt process $X^{(\delta,\lambda)}$ associated with it. For simplicity, we denote $X^{(\delta,\lambda)}$ by $X$. It follows from Theorem 3.1 and (4.6) that $X$ has quasi-continuous transition density function $q^{(\delta,\lambda)}(t, \xi, \eta)$ on $(0, \infty) \times G \times G$ with respect to the measure $\nu$.

For reader's convenience, let us summarize some notations here.

| | | |
|---|---|---|
| $p$ | heat kernel for $Y$, | corresponding Dirichlet form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ |
| $q^{(\delta)}$ | heat kernel for $Y^{(\delta)}$, | corresponding Dirichlet form $(W, \mathcal{D}(W))$ |
| $q^{(\delta,\lambda)}$ | heat kernel for $X = X^{(\delta,\lambda)}$, | corresponding Dirichlet form $(W^{(\delta,\lambda)}, \mathcal{D}(W))$ |

### 4.4 Heat kernel upper bound for the case $\gamma_1 = \gamma_2 = 0$

In this subsection, we derive an upper estimate the heat kernel $q^{(\delta)}(t, \xi, \eta)$ for the case $\gamma_1 = \gamma_2 = 0$.

We see from the second part of §4.1 that the process $X^{(\delta,\lambda)}$ can be constructed from $Y^{(\delta)}$ by suppressing jumps of size larger than $\lambda$ and so $q^{(\delta,\lambda)}(t,\xi,\eta)$ is well-defined on $(0,\infty)\times(G\setminus\delta^{-1}\mathcal{N})\times(G\setminus\delta^{-1}\mathcal{N})$. Note that there is a constant $c_1>0$, independent of $\delta>0$ and $\lambda>0$, such that

$$\|J_\lambda^{(\delta)}\|_\infty \le \frac{c_1}{\phi^{(\delta)}(\lambda)V^{(\delta)}(\lambda)}.$$

Thus by Lemma 4.1(ii), we have for every $t>0$ and $\xi,\eta\in G\setminus\delta^{-1}\mathcal{N}$,

$$q^{(\delta)}(t,\xi,\eta)\le q^{(\delta,\lambda)}(t,\xi,\eta)+t\|J_\lambda^{(\delta)}\|_\infty \le q^{(\delta,\lambda)}(t,\xi,\eta)+c_1\frac{t}{\phi^{(\delta)}(\lambda)V^{(\delta)}(\lambda)}. \tag{4.7}$$

We now derive an upper estimate for $q^{(\delta,\lambda)}(t,\xi,\eta)$ by using a result in [9].

**Lemma 4.3** *There exists a constant $c_1>0$, independent of $\delta>0$ such that*

$$q^{(\delta,\lambda)}(t,x,y)\le\ c_1\,\frac{V(\delta)\phi(\delta)\,t}{V(\delta\rho(x,y))\phi(\delta\rho(x,y))} \tag{4.8}$$

*for every $\delta>0$, $0<t\le 1$, $x,y\in G\setminus\delta^{-1}\mathcal{N}$ with $\rho(x,y)\ge 1$ and $\lambda=\frac{\beta_1}{3(d+\beta_1)}\rho(x,y)$.*

**Proof.** We will follow an idea in [2]. By (4.3), (4.6) and the $X^{(\delta)}$-version (that is, with $W_1^{(\delta,\lambda)}$ in place of $\mathcal{E}$) of Theorem 3.1, there are constants $c_1,c_2,c_3>0$, independent of $\delta>0$ and $\lambda\ge\frac{\beta_1}{3(d+\beta_1)}$, such that

$$\theta^{(\delta)}(\|u\|_{L^2(\delta^{-1}F)})\le c_2W_1(u,u)\le c_3W_1^{(\delta,\lambda)}(u,u)\qquad \text{for } u\in\mathcal{D}(W) \text{ with } \|u\|_{L^1(\delta^{-1}F)}=1,$$

where

$$\theta^{(\delta)}(r):=\frac{r}{\phi^{(\delta)}((V^{(\delta)})^{-1}(c_1r^{-1}))}.$$

Note that $(W^{(\delta,\lambda)},\mathcal{D}(W))$ is the Dirichlet form for the 1-subprocess of $X^{(\delta,\lambda)}$. By the proof of Theorem 3.2, there are constants $c_4,c_5>0$, independent of $\delta>0$ and $\lambda\ge\frac{\beta_1}{3(d+\beta_1)}$, such that

$$q^{(\delta,\lambda)}(t,x,y)\le\frac{c_5}{V^{(\delta)}((\phi^{(\delta)})^{-1}(c_4t))}\qquad \text{for } x,y\in G\setminus\delta^{-1}\mathcal{N} \text{ and } 0<t\le 1.$$

On the other hand, it follows from (1.2), (1.10) and (4.3) that there is a constant $c>0$, independent of $\delta>0$ and $\lambda>0$, such that

$$\frac{c_5}{V^{(\delta)}((\phi^{(\delta)})^{-1}(c_4t))}\le c\,t^{-d/\beta_1}\qquad \text{for } 0<t\le 1.$$

This together with Theorem 3.25 of [9] implies that there exist constants $C>0$ and $c>0$, independent of $\lambda>0$ and $\delta>0$ such that

$$q^{(\delta,\lambda)}(t,x,y)\le c\,t^{-d/\beta_1}\ \exp\left(-|\psi(y)-\psi(x)|+C\,\Gamma(\psi)^2\,t\right) \tag{4.9}$$

for all $t \in (0,1]$, $x, y \in G \setminus \delta^{-1}\mathcal{N}$ and every $\lambda > 0$, and for some $\psi$ satisfying $\Gamma(\psi) < \infty$, where

$$\Gamma(\psi)^2 = \|e^{-2\psi}\Gamma_\lambda[e^{\psi}]\|_\infty \vee \|e^{2\psi}\Gamma_\lambda[e^{-\psi}]\|_\infty,$$

where $\Gamma_\lambda$ is the *density* of $W^{(\delta,\lambda)}$ determined by

$$\Gamma_\lambda[v](\xi) = \int_{\rho(\eta,\xi)\le\lambda} (v(\eta) - v(\xi))^2 J^{(\delta)}(\eta,\xi)\nu(d\eta), \qquad \xi \in G. \tag{4.10}$$

Define

$$\mathcal{H}(\Gamma_\lambda) := \left\{ v : G \to \mathbb{R} \;\middle|\; \sup_{\xi\in G} \Gamma_\lambda[v](\xi) < \infty \right\}.$$

A key observation is that $\mathcal{H}(\Gamma_\lambda)$ contains the *cut-off distance function* $\psi$ given by

$$\psi(\xi) := \frac{s}{3}\left(\rho(\xi,x) \wedge \rho(x,y)\right) \qquad \text{for } \xi \in G, \tag{4.11}$$

where $s > 0$ is a number to be chosen later. Note that $|\psi(\eta) - \psi(\xi)| \le (s/3)\rho(\eta,\xi)$ for all $\xi, \eta \in G$. So

$$\begin{aligned}
e^{-2\psi(\xi)}\Gamma_\lambda[e^{\psi}](\xi) &= \int_{\rho(\eta,\xi)\le\lambda} (1 - e^{\psi(\eta)-\psi(\xi)})^2 J^{(\delta)}(\eta,\xi)\nu(d\eta) \\
&\le \int_{\rho(\eta,\xi)\le\lambda} (\psi(\eta) - \psi(\xi))^2\, e^{2|\psi(\eta)-\psi(\xi)|} J^{(\delta)}(\eta,\xi)\nu(d\eta) \\
&\le (\frac{s}{3})^2\, e^{2s\lambda/3} \int_{\rho(\eta,\xi)\le\lambda} \rho(\eta,\xi)^2 J^{(\delta)}(\eta,\xi)\nu(d\eta) \\
&\le cs^2 e^{2s\lambda/3} \int_0^\lambda \frac{t}{\phi^{(\delta)}(t)}\, dt \\
&\le cs^2 e^{2s\lambda/3} \frac{\lambda^2}{\phi^{(\delta)}(\lambda)} \\
&\le c\frac{e^{s\lambda}}{\phi^{(\delta)}(\lambda)},
\end{aligned}$$

for every $\xi \in F$, where the $\phi^{(\delta)}$-versions of Lemma 2.1(ii) and (1.14) are used in the third and fourth inequalities. The same estimate holds for $e^{2\psi(\xi)}\Gamma_\lambda[e^{-\psi}](\xi)$. Thus for $x, y \in G \setminus \delta^{-1}\mathcal{N}$ with $r := \rho(x,y) \ge 1$, the exponential part of (4.9) can be bounded from above by

$$-\frac{sr}{3} + ct\frac{e^{s\lambda}}{\phi^{(\delta)}(\lambda)}. \tag{4.12}$$

Now take $C = \frac{\beta_1}{3(d+\beta_1)}$, $\lambda = Cr$, and $s = \frac{1}{Cr}\log(\frac{\phi^{(\delta)}(r)}{t})$. Then, we can bound (4.12) as follows;

$$\log\left(\frac{t}{\phi^{(\delta)}(r)}\right)^{1/(3C)} + c\,\frac{\phi^{(\delta)}(r)}{t} \cdot \frac{t}{\phi^{(\delta)}(\lambda)} \le \log\left(\frac{t}{\phi^{(\delta)}(r)}\right)^{1/(3C)} + c,$$

where (1.13) is used in the last inequality. Putting this into (4.9), we obtain for $\xi, \eta \in G \setminus \delta^{-1}\mathcal{N}$,

$$q^{(\delta,\lambda)}(t,x,y) \le c\,t^{-d/\beta_1}\left(\frac{t}{\phi^{(\delta)}(r)}\right)^{\frac{d}{\beta_1}+1} = \frac{c\,t}{\phi^{(\delta)}(r)^{d/\beta_1}\phi^{(\delta)}(r)} \le \frac{c\,t}{V^{(\delta)}(r)\phi^{(\delta)}(r)}, \tag{4.13}$$

where in the last inequality, we used the fact $r \ge 1$, (1.11) and (1.13). This establishes (4.8). □

**Proof of Theorem 1.2 (upper bound estimate for the case of $\gamma_1 = \gamma_2 = 0$).**

Taking $\delta = 1$ and putting (4.8) into (4.7), we have, together with (3.5), that there is a constant $C > 0$, independent of $\delta > 0$ such that for every $t \le 1$ and $\xi, \eta \in G \setminus \delta^{-1}\mathcal{N}$ with $r := \rho(\xi,\eta) \ge 1$,

$$q^{(\delta)}(t,\xi,\eta) \le C\left(\frac{1}{V^{(\delta)}((\phi^{(\delta)})^{-1}(t))} \wedge \frac{t}{\phi^{(\delta)}(r)V^{(\delta)}(r)}\right). \tag{4.14}$$

This in particular implies, by taking $\delta = 1$, that for every $t \le 1$ and $x, y \in F\setminus\mathcal{N}$ with $r := \rho(x,y) \ge 1$,

$$p(t,x,y) \le C\left(\frac{1}{V(\phi^{-1}(t))} \wedge \frac{t}{\phi(r)V(r)}\right). \tag{4.15}$$

Thanks to (3.5), (4.15) holds for $t \ge \phi(\rho(x,y))$ as well, if suffices to consider the case $t < \phi(\rho(x,y))$ with either $r := \rho(x,y) < 1$ or $t > 1$. In either case, take $\delta > 0$ so that $c_5 \le \delta^{-1}r \le 2c_5$, where $c_5 \ge 1$ is the constant in (1.15). Then $\rho(\delta^{-1}x,\, \delta^{-1}y) \ge 1$ and $t/\phi(\delta) \asymp t/\phi(r) \le 1$. Thus by (4.4) and (4.14), for every $t > 0$,

$$\begin{aligned}
p(t,x,y) &= V(\delta)^{-1}\, q^{(\delta)}(t/\phi(\delta),\, \delta^{-1}x, \delta^{-1}y) \\
&\le C\,V(\delta)^{-1}\left(\frac{1}{V^{(\delta)}((\phi^{(\delta)})^{-1}(t/\phi(\delta))} \wedge \frac{t/\phi(\delta)}{\phi^{(\delta)}(\rho(\delta^{-1}x,\delta^{-1}y))\,V^{(\delta)}(\rho(\delta^{-1}x,\delta^{-1}y))}\right) \\
&= C\left(\frac{1}{V(\phi^{-1}(t))} \wedge \frac{t}{\phi(\rho(x,y))\,V(\rho(x,y))}\right).
\end{aligned}$$

where in the last inequality we used (4.3) and assumptions (1.11)-(1.12) and (1.15), which ensures that the new constant $C > 0$ in the last inequality is independent of $t > 0$ and $x, y \in F \setminus \mathcal{N}$. This establishes the heat kernel upper bound estimate in Theorem 1.2 for the case of $\gamma_1 = \gamma_2 = 0$. □

**Remark 4.4** We used the scaling argument in the above proof of Theorem 1.2 for the case of $\gamma_1 = \gamma_2 = 0$ to circumvent the restriction that the conclusion of Lemma 4.3 holds for $\rho(x,y) \ge 1$. As we noted above, the condition $\rho(x,y) \ge 1$ of Lemma 4.3 is only used in the last inequality in (4.13). When $d_0 = d$ and $\beta_1 = \beta_2$, that last inequality holds for every $r := \rho(x,y) > 0$. Moreover, in this case, the Nash's inequality implies that the upper bound in (3.5) is $c\,t^{-d/\beta_1}$ for every $t > 0$ and so (4.9) holds for every $t > 0$. Thus in the case of $d_0 = d$ and $\beta_1 = \beta_2$, taking $\delta = 1$, we have (4.8) holds for every $t > 0$, $\delta = 1$, and every pair of $x, y \in F \setminus \mathcal{N}$ having $\rho(x,y) > 0$ and $\lambda = \frac{\beta_1}{3(d+\beta_1)}\rho(x,y)$. This together with (4.7) of $\delta = 1$ and (3.5) establishes the upper bound estimate in Theorem 1.2, without using the scaling argument. Thus in the case of $d_0 = d$ and $\beta_1 = \beta_2$, we do not need to assume (1.15) and the heat kernel estimate in [11, 16] can be obtained without using a scaling argument.

### 4.5 Heat kernel upper bound for the cases $\gamma_2 \geq \gamma_1 > 0$

In this subsection, we will obtain the upper estimate the heat kernel $p(t,x,y)$ of $Y$ for the case $\gamma_2 \geq \gamma_1 > 0$. Let $\phi_1$ be the function in Remark 1.1 related to $\phi$. Let $Y$ and $Y^*$ be the Hunt process associated with the Dirichlet form in Section 2 corresponding to the jumping intensity kernels

$$J(x,y) = \frac{c(x,y)}{V(\rho(x,y))\phi(\rho(x,y))} \qquad \text{and} \qquad J_*(x,y) = \frac{c(x,y)}{V(\rho(x,y))\phi_1(\rho(x,y))},$$

respectively. Here $c(x,y)$ is a symmetric function bounded between two positive constants. Clearly, $J(x,y) \leq J_*(x,y)$ and

$$\mathcal{J}(x) := \int_F (J_*(x,y) - J(x,y))\mu(dy) \leq c_3 \qquad \text{for all } x \in F.$$

We know from Section 3 that there are quasi-continuous transition density functions $p(t,x,y)$ and $p^*(t,x,y)$ on $(0,\infty) \times F \times F$ for $Y$ and $Y^*$, respectively. By the first part of §4.1, process $Y^*$ can be obtained from $Y$ by adding more jumps and so $p^*(t,x,y)$ is well-defined on $(0,\infty) \times (F \setminus \mathcal{N}) \times (F \setminus \mathcal{N})$, where $\mathcal{N}$ is the properly exceptional set of $Y$ in Theorem 3.2.

**Lemma 4.5** *For every $t_0 > 0$, there is a constant $C = C(t_0) > 0$ such that*

$$p(t,x,y) \leq C\left(\frac{1}{V(\phi^{-1}(t))} \wedge \frac{t}{V(\rho(x,y))\phi_1(\rho(x,y))}\right),$$

*for every $t \in (0, t_0]$ and every $x, y \in F \setminus \mathcal{N}$, where $\phi_1^{-1}$ is a inverse function of $\phi_1$.*

**Proof.** Using Lemma 4.1(i), we have for q.e. $x \in F$ and every positive Borel measurable function $f$ on $F$,

$$\mathbb{E}_x\left[f(Y_t)\right] \leq e^{t\|\mathcal{J}\|_\infty}\, \mathbb{E}_x\left[f(Y_t);\, Y_s = Y_s^* \text{ for all } 0 \leq s \leq t\right] \leq e^{t\|\mathcal{J}\|_\infty}\, \mathbb{E}_x\left[f(Y_t^*)\right].$$

This implies that

$$p(t,x,y) \leq e^{t\|\mathcal{J}\|_\infty}\, p^*(t,x,y) \qquad \text{for q.e. } x, y \in F.$$

Note that for $t \in (0, t_0]$, $\phi^{-1}(t) \approx \phi_1^{-1}(t)$. The conclusion of this lemma now follows from upper bound estimate in Theorem 1.2 for $p^*(t,x,y)$ (of the case $\gamma_1 = \gamma_2 = 0$), which is proved in the last subsection. □

By this lemma, we see that the upper bound of Theorem 1.2 holds for $r := \rho(x,y) \le 1$ when $\gamma_2 \ge \gamma_1 > 0$. In the remaining of this subsection, we assume $\gamma_2 \ge \gamma_1 > 0$ and $\rho(x,y) > 1$. Recall the notation of $q^{(\delta,\lambda)}$ from §4.3.

**Lemma 4.6** *There exist $c_1, c_2 > 0$ such that*

$$q^{(1,\lambda)}(t,x,y) \le \; c_1 \frac{t}{V(r)\phi_1(r)} \; \exp\left(-c_2\gamma_1 r\right) \tag{4.16}$$

*for all $t \in (0,1]$, $x, y \in F \setminus \mathcal{N}$ with $r := \rho(x,y) > 0$, and $\lambda = \frac{\beta_1}{8(d+\beta_1)} r$.*

**Proof.** In view of Lemma 4.5, it suffices to prove the Lemma for $r \ge r_0$, where $r_0 \ge 1$ is a constant to be chosen later.

The idea of the proof is similar to that for Lemma 4.3, but some modification is needed in order to optimize (4.9). Let $\Gamma_\lambda$ be as in the proof Lemma 4.3 but with $\delta = 1$. Take the cut-off function $\psi$ to be

$$\psi(\xi) := \frac{s+\gamma_1}{3} \left(\rho(\xi,x) \wedge \rho(x,y)\right) \qquad \text{for } \xi \in F, \tag{4.17}$$

where $\gamma_1$ is in (1.12) and $s > 0$ is a constant to be chosen later. Note that $|\psi(\eta) - \psi(\xi)| \le \frac{s+\gamma_1}{3}\rho(\eta,\xi)$ for all $\xi, \eta \in F$. So, using the fact $\gamma_1 > 0$, we have

$$\begin{aligned}
e^{-2\psi(\xi)}\Gamma_\lambda[e^\psi](\xi) &= \int_{\rho(\eta,\xi)\le\lambda} (1 - e^{\psi(\eta)-\psi(\xi)})^2 J(\eta,\xi)\mu(d\eta) \\
&\le \int_{\rho(\eta,\xi)\le\lambda} (\psi(\eta) - \psi(\xi))^2 \; e^{2|\psi(\eta)-\psi(\xi)|} J(\eta,\xi)\mu(d\eta) \\
&\le \left(\frac{s+\gamma_1}{3}\right)^2 \; e^{2s\lambda/3} \int_{\rho(\eta,\xi)\le\lambda} \rho(\eta,\xi)^2 e^{2\gamma_1\rho(\eta,\xi)/3} J(\eta,\xi)\mu(d\eta) \\
&\le \frac{c}{3}\,(s+\gamma_1)^2\, e^{2s\lambda/3} \int_0^\lambda \frac{t e^{2\gamma_1 t/3}}{\phi(t)}\, dt \\
&\le \frac{c}{3}\,(s+\gamma_1)^2\, e^{2s\lambda/3} \left(c_1 + c_2 \int_1^\lambda e^{-c_3\gamma_1 t} dt\right) \\
&\le \frac{c}{3}\,(s+\gamma_1)^2\, e^{2s\lambda/3},
\end{aligned}$$

for every $\xi \in F$, where the third inequality can be proved by straightforwardly modifying the proof of Lemma 2.1(ii), and the lower bound of (1.12) is used in the second to the last inequality. The same estimate holds for $e^{2\psi(\xi)}\Gamma_\lambda[e^{-\psi}](\xi)$. Thus the exponential part of (4.9) can be bounded from above by

$$\frac{s+\gamma_1}{3}\left(-r + c\,t\,(s+\gamma_1)\, e^{2s\lambda/3}\right). \tag{4.18}$$

Now take $C = \frac{\beta_1}{8(d+\beta_1)}$, $\lambda = Cr$, and $s = \frac{3}{4Cr}\log(\frac{ar}{t})$. Here $a \in (0,1]$ is a constant such that

$$\frac{3ac}{4C} \sup_{0<s\le 1} \sqrt{s}\log\frac{1}{s} + \sqrt{a}c\gamma_1 < 1/2,$$

where $c > 0$ is the constant in (4.18). Take $r_0 = 1/a$. Then for $r \geq r_0$ and $t \in (0,\, 1]$,

$$\begin{aligned}
& \frac{s+\gamma_1}{3}\left(-r + ct(s+\gamma_1)e^{\frac{1}{2}\log\frac{ar}{t}}\right) \\
= \; & \frac{s+\gamma_1}{3}\left(-r + c\,(s+\gamma_1)\,(art)^{1/2}\right) \\
= \; & \frac{s+\gamma_1}{3}\left(-r + \frac{3ac}{4C}\sqrt{\frac{t}{ar}}\log\frac{ar}{t} + a^{1/2}c\gamma_1 r\left(\frac{t}{r}\right)^{1/2}\right) \\
\leq \; & -r(s+\gamma_1)/6 \\
= \; & -\frac{d+\beta_1}{\beta_1}\log\frac{ar}{t} - \frac{r\gamma_1}{6}.
\end{aligned}$$

Putting this into (4.9), we obtain

$$q^{(1,\lambda)}(t,x,y) \leq c\,t^{-d/\beta_1}\;\exp\left(-r(s+\gamma_1)/6\right) = c\,t^{-d/\beta_1}\left(\frac{t}{ar}\right)^{\frac{d}{\beta_1}+1} e^{-\frac{\gamma_1 r}{6}} \leq \frac{c_1\,t}{V(r)\phi_1(r)}e^{-c_2\gamma_1 r},$$

where in the last inequality, we used the fact $r \geq r_0 \geq 1$. This proves (4.16). □

**Proof of Theorem 1.2 (upper bound estimate for the case of $\gamma_2 > \gamma_1 > 0$).** Note that by (1.4) and (1.7)-(1.8), there are constants $c_1, c_2 > 0$ such that for every $\lambda > 0$ and $\xi \in F$,

$$\left\|\mathbf{1}_{\rho(\xi,\cdot)>\lambda}J(\xi,\cdot)\right\|_\infty \leq \frac{c_1}{V(\lambda)\phi(c_2\lambda)}.$$

It then follows from (4.16) and (4.7) with $\delta = 1$ that for every $t \in (0,1]$ and $x, y \in F \setminus \mathcal{N}$ (taking $\lambda = \frac{\beta_1}{8(d+\beta_1)}\rho(x,y)$ and noting the doubling property (1.7) of $V$), we have

$$p(t,x,y) \leq q^{(1,\lambda)}(t,x,y) + t\sup_{\xi\in F}\left\|\mathbf{1}_{\rho(\xi,\cdot)>\lambda}J(\xi,\cdot)\right\|_\infty \leq \frac{c_3\,t}{V(\rho(x,y))\phi(c_4\rho(x,y))}$$

This together with (4.14) yields the desired upper bound estimate for $p(t,x,y)$ for $t \in (0,1]$ and $x, y \in F \setminus \mathcal{N}$. □

## 4.6 Conservativeness

Let $Y$ be the Hunt process associated with the regular Dirichlet form $(\mathcal{E}, \mathcal{F})$ defined by (1.5), where the jumping intensity kernel is specified by (1.9). As mentioned in the first paragraph of §4.2, only conditions (1.1)-(1.3) and (1.7)-(1.9) are assumed in this subsection. In particular, we do *not* need the condition (1.15) nor assumption $\mu(F) = \infty$ in this subsection.

**Theorem 4.7** *The process $Y$ is conservative; that is, $Y$ has infinite lifetime.*

**Proof.** Let $\overline{Y}$ be the Hunt process associated with the Dirichlet form on $F$ given by (1.5) but with jumping kernel $J_0(x,y) := J(x,y)1_{\{d(x,y)\le 1\}}$ in place of $J$ there and with $\mathcal{F} = \overline{\mathcal{D}(\mathcal{E})}^{\mathcal{E}_1}$. Clearly $J_0(x,y) \le J(x,y)$ and by Lemma 2.1(i),

$$\mathcal{J}(x) := \int_F (J(x,y) - J_0(x,y))\mu(dy) = \int_{\{y\in F:\, d(x,y)>1\}} J(x,y)\mu(dy) \le c_0 \qquad \text{for every } x \in F.$$

Thus it suffices to show that $\overline{Y}$ is conservative. This is because the process $Y$ can be obtained from $\overline{Y}$ through Meyer's construction as discussed in the first paragraph of Section 4.1, and therefore the conservativeness of $Y$ follows immediately from that of $\overline{Y}$.

To show that $\overline{Y}$ is conservative, we look at reflected jump processes with jumping kernel $J_0$ in big balls. Let $x_0$, $\kappa$ and $r_n \ge 100$ be as in condition (1.1). Define $B_n = \overline{B(x_0, r_n)}$ and

$$\begin{aligned}\mathcal{E}^{(n)}(f,f) &= \int_{B_n}\int_{B_n} (f(x)-f(y))^2 J_0(x,y)\mu(dx)\mu(dy),\\ \mathcal{F}^{(n)} &= \overline{\{f \in C(B_n) : \mathcal{E}^{(n)}(f,f) < \infty\}}^{\mathcal{E}_1^{(n)}},\end{aligned}$$

where $\mathcal{E}_1^{(n)}(u,u) := \mathcal{E}^{(n)}(u,u) + \int_{B_n} u(x)^2\mu(dx)$. Clearly $(\mathcal{E}^{(n)}, \mathcal{F}^{(n)})$ is a regular symmetric Dirichlet form on $L^2(B_n,\mu)$. Let $Y^{(n)}$ be the Hunt process on $B_n$ associated with $(\mathcal{E}^{(n)}, \mathcal{F}^{(n)})$. Since constant function $1 \in \mathcal{F}^{(n)}$ with $\mathcal{E}^{(n)}(1,1) = 0$, $Y^{(n)}$ is recurrent and so $Y^{(n)}$ is conservative. We claim that there is a constant $c_2 > 0$ such that

$$\theta(\|u\|_2^2) \le c_2\,(\mathcal{E}^{(n)}(u,u) + \|u\|_2^2), \qquad \text{for all } u \in \mathcal{F}^{(n)} \text{ with } \|u\|_1 = 1, \tag{4.19}$$

where $\theta(r) = \frac{r}{\phi(V^{-1}(c_1 r^{-1}))}$ as in Theorem 3.1 with constant $c_1 > 0$ independent of $n \ge 1$. Here and in the remaining proof of this theorem, positive constants $c_i$'s, unless otherwise specified, are independent of $n \ge 1$. Indeed, in view of the condition (1.1) and (1.7)-(1.8), there are universal constants $c_3, c_4, c_5 > 0$ such that for every $n \ge 1$, $x \in B_n$ and $0 < r \le 1$,

$$c_3 V(r) \ge \mu(B(x,r)\cap B_n) \ge c_4 V(\kappa r) \ge c_5 V(r).$$

Thus by the same argument as that leads to (3.3), we have

$$\|u\|_2^2 \le c_6\left(\phi(r)\mathcal{E}^{(n)}(u,u) + \frac{1}{V(r)}\right) \le c_6\left(\phi(r)\left(\mathcal{E}^{(n)}(u,u) + \|u\|_2^2\right) + \frac{1}{V(r)}\right).$$

Note that $r \mapsto (\phi(r)V(r))^{-1}$ is a strictly decreasing function on $(0,1]$ with $\lim_{r\to 0+}(\phi(r)V(r))^{-1} = \infty$. If there exists $r_0 \le 1$ such that

$$(\phi(r_0)V(r_0))^{-1} = \mathcal{E}^{(n)}(u,u) + \|u\|_2^2,$$

then we obtain (4.19) by the same argument as in the proof of Theorem 3.1. Otherwise we have $L := (\phi(1)V(1))^{-1} \ge \|u\|_2^2$. On the other hand, by definition of $\theta$, we know $\theta(x) \le c_L x$ for $x \le L$, so

(4.19) holds in this case as well. Observe that $(\mathcal{E}_1^{(n)}, \mathcal{F}^{(n)})$ is the Dirichlet form of the 1-subprocess of $Y^{(n)}$. Thus by (4.19), Theorem 3.2 and [13], there is a properly exceptional set $\mathcal{N}_n$ of $Y^{(n)}$, a positive symmetric kernel $p^{(n)}(t,x,y)$ defined on $(0,\infty) \times (B_n \setminus \mathcal{N}_n) \times (B_n \setminus \mathcal{N}_n)$ such that for every $x \in B_n \setminus \mathcal{N}_n$, $Y^{(n)}$ has infinite lifetime $\mathbb{P}^x$-a.s., $\mathbb{P}^x(Y_t^{(n)} \in dy) = p^{(n)}(t,x,y)\mu(dy)$ on $B_n$ and

$$p^{(n)}(t,x,y) \le \frac{c_7\, e^t}{V(\phi^{-1}(c_1 t))} \qquad \text{for every } n \ge 1,\ x,y \in B_n \setminus \mathcal{N}_n \text{ and for every } t>0. \tag{4.20}$$

Moreover, for every $t>0$ and $y \in B_n \setminus \mathcal{N}_n$, $x \mapsto p^{(n)}(t,x,y)$ is $\mathcal{E}^{(n)}$-quasi-continuous on $B_n$. Then, the similar proof as that of Lemma 4.3 up to (4.12) (with $s=\delta=\lambda=1$) gives

$$p^{(n)}(t,x,y) \le c_8 t^{-d/\beta_1} \exp(-c_9 \rho(x,y)) \qquad \text{for every } x,y \in B_n \setminus \mathcal{N}_n \text{ and for every } t \in (0,2]. \tag{4.21}$$

(Note that only conditions (1.2)-(1.3) and (1.7)-(1.9) are needed for the proof of the above inequality.) So for $x \in B_n \setminus \mathcal{N}_n$, $t \in [1,2]$ and $r \in (0, r_n)$,

$$\begin{aligned} \mathbb{P}^x\left(\rho(Y_t^{(n)}, x) \ge r\right) &= \int_{B_n \setminus B(x,r)} p^{(n)}(t,x,y)\mu(dy) \\ &\le c_8 \int_{B_n \setminus B(x,r)} e^{-c_9 \rho(x,y)} \mu(dy) \\ &\le c_{10} e^{-c_{11} r}, \end{aligned} \tag{4.22}$$

where we used the fact $Y^{(n)}$ that is conservative in the first equality and (4.21) in the first inequality. Let $I_r := c_{10} e^{-c_{11} r}$ and define $\sigma_r := \inf\{t \ge 0 : \rho(Y_t^{(n)}, Y_0^{(n)}) > r\}$. Then by the conservativeness of $Y^{(n)}$, (4.22) and the strong Markov property of $Y^{(n)}$, for every $x \in B_n \setminus \mathcal{N}_n$ and $r \in (0, r_n)$,

$$\begin{aligned} \mathbb{P}^x\left(\sup_{u \le 1} \rho(Y_u^{(n)}, Y_0^{(n)}) > r\right) &= \mathbb{P}^x(\sigma_r < 1) \\ &\le \mathbb{P}^x\left(\sigma_r < 1 \text{ and } \rho(Y_2^{(n)}, x) \le r/2\right) + \mathbb{P}^x\left(\rho(Y_2^{(n)}, x) > r/2\right) \\ &\le \mathbb{P}^x\left(\sigma_r < 1 \text{ and } \rho(Y_2^{(n)}, Y_{\sigma_r}^{(n)}) > r/2\right) + I_{r/2} \\ &\le \mathbb{E}^x\left[\mathbf{1}_{\{T:=\sigma_r<1\}} \mathbb{P}^{Y_T^{(n)}}\left(\rho(Y_{2-T}^{(n)}, Y_0^{(n)}) \ge r/2\right)\right] + I_{r/2} \\ &\le \sup_{y \in (B_n \setminus B(x,r)) \setminus \mathcal{N}_R} \sup_{u \le 1} \mathbb{P}^y\left(\rho(Y_{2-u}^{(n)}, y) \ge r/2\right) + I_{r/2} \\ &\le 2I_{r/2} = 2c_{10} e^{-c_{11} r/2}. \end{aligned} \tag{4.23}$$

Note that $\mu(\mathcal{N}_n) = 0$ and for $x \in B_{r_n - 1} \setminus \mathcal{N}_n$, $Y^{(n)}$ has the same distribution as that of $\overline{Y}$ before $Y^{(n)}$ leaves the ball $B_{r_n-1}$. For every $r_0 > 0$, take $K$ to be an integer that is larger than $2r_0 + 1$. We have in particular from the above display that for $\mu$-a.e. $x \in B_{r_0}$,

$$\mathbb{P}^x\left(\sup_{u \le 1} \rho(Y_u^{(n)}, Y_0^{(n)}) \le r\right) \ge 1 - 2c_{10} e^{-c_{11} r/2} \qquad \text{for every } n \text{ with } r_n > K \text{ and for } r < r_n - 1.$$

Let $\{\overline{P}_t, t \geq 0\}$ denote the transition semigroup of $\overline{Y}$. Thus we have for $\mu$-a.e. $x \in B_{r_0}$,

$$\mathbb{P}^x \left( \overline{\zeta} > 1 \text{ and } \sup_{u \leq 1} \rho(\overline{Y}_u, \overline{Y}_0) \leq r \right) \geq 1 - 2c_{10} e^{-c_{11} r/2} \qquad \text{for every } r > 0.$$

It follows then for $\mu$-a.e. $x \in B_{r_0}$,

$$\mathbb{P}^x(\overline{Y}_1 \in F) = 1.$$

Let $r_0$ increase to infinity through an increasing sequence of positive integers. We have

$$\mathbb{P}^x(\overline{Y}_1 \in F) = 1 \qquad \text{for } \mu\text{-a.e. } x \in F$$

and consequently, by the Markov property of $\overline{Y}$, we have for $\mu$-a.e. $x \in F$,

$$\mathbb{P}^x(\overline{Y}_t \in F) = 1 \qquad \text{for every rational } t > 0.$$

Since for each rational $t > 0$, $\overline{P}_t 1$ is finely continuous and $\overline{P}_t 1 = 1$ $\mu$-a.e. on $F$, we must have $\overline{P}_t 1 = 1$ q.e. on $F$. Let $\overline{\zeta}$ denote the lifetime of $\overline{Y}$. The above says that $\mathbb{P}^x(\overline{\zeta} = \infty) = 1$ for q.e. $x \in F$. By Meyer's construction of $Y$ from $\overline{Y}$, we have proved that the process $Y$ is conservative. $\square$

## 4.7 Tightness and some lower bound estimate

We first give a well-known formula on Lévy system of $Y$. The Hunt process $Y$ on $F \setminus \mathcal{N}$ has a Lévy system, which is closely related to the jumping measure $J(\xi, \eta)\mu(d\xi)\mu(d\eta)$ (cf. [14]). The following result corresponds to Lemma 4.7 in [11]. See Appendix A below for more details.

**Lemma 4.8** *Let $f$ be a non-negative measurable function on $\mathbb{R}_+ \times F \times F$ that vanishes along the diagonal. Then for every $t \geq 0$, $x \in F \setminus \mathcal{N}$ and stopping time $T$ (with respect to the filtration of $Y$),*

$$\mathbb{E}^x \left[ \sum_{s \leq T} f(s, Y_{s-}, Y_s) \right] = \mathbb{E}^x \left[ \int_0^T \left( \int_F f(s, Y_s, y) J(Y_s, y) \mu(dy) \right) ds \right].$$

For $A \subset F$, let $\tau_A := \inf\{t \geq 0 : Y_t \notin A\}$. The following proposition corresponds to Proposition 4.1 in [11], which, together with the parabolic Harnack inequality, will be used to get the lower bound heat kernel estimate for $p(t, x, y)$.

**Proposition 4.9** *For each $A > 0$ and $0 < B < 1$, there exists $\gamma = \gamma(A, B) \in (0, 1/2)$ such that for every $r \in (0, 1]$ (resp. $r > 0$ when $\gamma_1 = \gamma_2 = 0$) and $x \in F \setminus \mathcal{N}$,*

$$\mathbb{P}^x \left( \tau_{B(x, Ar)} < \gamma\, \phi(r) \right) \leq B.$$

**Proof.** Let $x \in F \setminus \mathcal{N}$. By the upper bound estimate in Theorem 1.2, for every $s > 0$ and $t \leq 1$ (resp. for every $t > 0$ in the case of $\gamma_1 = \gamma_2 = 0$)

$$\begin{aligned}
\mathbb{P}^x\left(\rho(Y_t, x) \geq s\right) &= \int_{B(x,s)^c} p(t,x,y)\mu(dy) + \mathbb{P}^x(\zeta \leq s) \qquad (4.24)\\
&\leq \int_{B(x,s)^c} \frac{c\,t}{V(\rho(x,y))\phi(c_1\rho(x,y))}\mu(dy)\\
&\leq \frac{ct}{\phi(c_1 s)}, \qquad (4.25)
\end{aligned}$$

where we used the conservativeness of $Y$ in the first equality, and Lemma 2.1(i) in the last inequality. Let $I_{s,t} := \frac{ct}{\phi(c_1 s)}$ and define $\sigma_s := \inf\{t \geq 0 : \rho(Y_t, Y_0) > s\}$. Then by (4.25) and the strong Markov property of $Y$, we can compute similarly to (4.23) and obtain

$$\mathbb{P}^x\left(\sup_{u \leq t} \rho(Y_u, Y_0) > s\right) \leq 2I_{s/2,2t} = \frac{c\,t}{\phi(c_1 s/2)}, \qquad (4.26)$$

for every $t \leq 1/2$ and $s > 0$. Recall that $\phi$ is an increasing function with $\phi(0) = 0$ and $\phi(1) = 1$. The above implies that for every $x \in F \setminus \mathcal{N}$ and $r \in (0,1]$ (resp. for every $r > 0$ in the case of $\gamma_1 = \gamma_2 = 0$),

$$\mathbb{P}^x\left(\sup_{u \leq \phi(r)\,t} \rho(Y_u, x) > r\,s\right) \leq \frac{c\,\phi(r)\,t}{\phi(c_1 r s/2)} \qquad \text{for every } s > 0 \text{ and } t \leq 1/2. \qquad (4.27)$$

For $A \geq 2/c_1$ and $B \in (0,\,1)$, by (1.2) we can choose $t_0 < 1/2$ so that

$$\frac{c\phi(r)t_0}{\phi(c_1 r A/2)} \leq \frac{ct_0}{A^{\beta_1}} < B \qquad \text{for every } r > 0. \qquad (4.28)$$

Thus we have by (4.27),

$$\mathbb{P}^x\left(\sup_{s \leq \phi(r)\,t_0} \rho(Y_s, x) \geq Ar\right) \leq B \qquad \text{for every } r \in (0,1] \;\; (\text{resp. } r > 0 \text{ when } \gamma_1 = \gamma_2 = 0).$$

This proves the Proposition with $\gamma = t_0$ for $A \geq 2/c_1$ and $B \in (0,\,1)$.

Now for $A \in (0, 2/c_1)$ and $B \in (0,\,1)$, let $\widetilde{t}_0$ be as in (4.28) corresponding to $A = 2/c_1$ and $B$. As we can write $Ar = \frac{2}{c_1}\frac{c_1 Ar}{2}$ with $\frac{c_1 Ar}{2} < r$, we have from the above

$$\mathbb{P}^x\left(\sup_{s \leq \phi(c_1 Ar/2)\,\widetilde{t}_0} \rho(Y_s, x) \geq Ar\right) \leq B \qquad \text{for every } r \in (0,1] \;\; (\text{resp. } r > 0 \text{ when } \gamma_1 = \gamma_2 = 0).$$

We have from (1.12) and (1.13) that there is a constant $c_2 > 0$ such that

$$\phi(c_1 Ar/2) \geq c_2 \phi(r) \qquad \text{for every } r \in (0,1] \;\; (\text{resp. } r > 0 \text{ when } \gamma_1 = \gamma_2 = 0).$$

It follows then

$$\mathbb{P}^x\left(\sup_{s\le c_2\phi(r)\,\widetilde{t}_0}\rho(Y_s,x)\ge Ar\right)\le B\qquad\text{for every } r\in(0,1]\ \ (\text{resp. } r>0 \text{ when } \gamma_1=\gamma_2=0).$$

This proves the Proposition with $\gamma=\min\left\{c_2\widetilde{t}_0,\,1/3\right\}$ for $A\in(0,2/c_1)$ and $B\in(0,\,1)$. $\square$

**Remark 4.10** (1) There is a gap in [11, 16] in the proof corresponding to that of this proposition. The conservativeness of $Y$ is not addressed in [11, 16]. Without the conservativeness of $Y$, just as in (4.24) above, there should be an $\mathbb{P}^x(\zeta\le t)$ term in [11, (4.14)] and in [16, (4.18)]. Theorem 4.7 of this paper has fixed this gap.

(2) We point out that, in the case of the metric measure space $F$ being a Euclidean closed subset, using the fact that $Y$ is conservative and the theory of reflected Dirichlet forms (cf. [10]), it can be shown that the domain $\mathcal{F}$ of the Dirichlet form of $Y$ defined in Proposition 2.2 is in fact equal to

$$\mathcal{F}=\{u\in L^2(F,\mu):\mathcal{E}(u,u)<\infty\}.$$

However since this result is not needed in this paper, we omit its proof. $\square$

In the following, we will denote $\gamma(1/2,1/2)$ in Proposition 4.9 by $\gamma$.

**Proposition 4.11** *There exist constants $c_1\ge 2, c_2>0$ and $c_3>0$ such that the following holds for every $\xi,\eta\in F\setminus\mathcal{N}$ with $\rho(\xi,\eta)\ge c_1\phi^{-1}(t)$ and for every $t\in(0,1]$ (resp. for every $t>0$ in the case of $\gamma_1=\gamma_2=0$),*

$$\mathbb{P}^\xi(Y_t\in B(\eta,c_1\phi^{-1}(t)))\ge c_2\frac{tV(\phi^{-1}(t))}{V(\rho(\eta,\xi))\phi(c_3\rho(\eta,\xi))}. \tag{4.29}$$

**Proof.** By Proposition 4.9, starting at $z\in B(y,\phi^{-1}(t))$, with probability at least $1/2$ the process $Y$ does not move more than $\phi^{-1}(\gamma^{-1}t)/2$ by time $t$. By (1.12) and (1.13), there is a constant $c_0>0$ such that

$$\tfrac{1}{2}\phi^{-1}(\gamma^{-1}t)\le c_0\phi^{-1}(t)\qquad\text{for every } t\in(0,1]\ \ (\text{resp. for every } t>0 \text{ when } \gamma_1=\gamma_2=0).$$

Thus, it is sufficient to show for some constant $c_4>0$,

$$\mathbb{P}^x\left(Y\text{ hits ball }B(y,\phi^{-1}(t))\text{ before }t\right)\ge c_2\frac{V(\phi^{-1}(t))}{V(\rho(\eta,\xi))}\cdot\frac{t}{\phi(c_4\rho(\eta,\xi))} \tag{4.30}$$

for all $\rho(x,y) \geq 3c_0\phi^{-1}(t)$ and $t \in (0,1]$ (resp. $t > 0$ when $\gamma_1 = \gamma_2 = 0$). Now with $B_x := B(x,\, c_0\phi^{-1}(t))$, $B_y := B(y,\, c_0\phi^{-1}(t))$ and $\tau_x := \tau_{B_x}$, it follows from Proposition 4.9,

$$\mathbb{E}^x\left[t \wedge \tau_x\right] \geq \gamma t\, \mathbb{P}^x\left(\tau_x \geq \gamma t\right) \geq t/2, \tag{4.31}$$

for each $t \leq 1$ (resp. $t > 0$ when $\gamma_1 = \gamma_2 = 0$). Thus, from Lemma 4.8,

$$\begin{aligned}
& \mathbb{P}^x\left(Y \text{ hits ball } B(y,\, \phi^{-1}(t)) \text{ by time } t\right) \\
\geq\ & \mathbb{P}^x(Y_{t\wedge\tau_x} \in B(y,\phi^{-1}(t)) \text{ and } t\wedge\tau_x \text{ is a jumping time }) \\
\geq\ & \mathbb{E}^x\left[\int_0^{t\wedge\tau_x}\int_{B_y}\frac{c}{V(\rho(Y_s,u))\phi(\rho(c_2Y_s,u))}\mu(du)ds\right] \\
\geq\ & c_3\mathbb{E}^x\left[t\wedge\tau_x\right]\int_{B_y}\frac{1}{V(\rho(x,y))\phi(c_4\rho(x,y))}\,\mu(du) \\
\geq\ & \frac{c_3\, t}{2}\frac{\mu(B(y,c_0\phi^{-1}(t)))}{V(\rho(x,y))\phi(c_4\rho(x,y))} \\
\geq\ & c_5\frac{tV(\phi^{-1}(t))}{V(\rho(x,y))\phi(c_4\rho(x,y))}.
\end{aligned}$$

Here in the fourth inequality, (4.31) is used. This establishes (4.30). □

## 4.8 Parabolic Harnack inequality

In [11], the parabolic Harnack inequality (Proposition 4.3 in [11]) is proved in order to obtain the upper and lower heat kernel estimates. Here we will introduce the parabolic Harnack inequality and will use it to obtain the heat kernel lower bound. Note that the heat kernel upper bound is already proved without the parabolic Harnack inequality.

We first introduce a space-time process $Z_s := (V_s, Y_s)$, where $V_s = V_0 + s$. The filtration generated by $Z$ satisfying the usual condition will be denoted as $\{\widetilde{\mathcal{F}}_s;\, s \geq 0\}$. The law of the space-time process $s \mapsto Z_s$ starting from $(t,x)$ will be denoted as $\mathbb{P}^{(t,x)}$.

We say that a non-negative Borel measurable function $h(t,x)$ on $[0,\infty)\times G$ is *parabolic* in a relatively open subset $D$ of $[0,\infty)\times G$ if for every relatively compact open subset $D_1$ of $D$, $h(t,x) = \mathbb{E}^{(t,x)}[h(Z_{\tau_{D_1}})]$ for every $(t,x) \in D_1$, where $Z_s = (V_s, Y_s)$ is the space-time process with $V_s = V_0 + s$ and $\tau_{D_1} = \inf\{s > 0 :\, Z_s \notin D_1\}$.

For each $r, t > 0$, we define

$$Q(t,x,r) := [t, t+\gamma\phi(r)] \times B(x,r).$$

**Theorem 4.12** *For every $0 < \delta \le \gamma$, there exists $c_1 > 0$ such that for every $z \in F$, $R \in (0,1]$ (resp. $R > 0$ when $\gamma_1 = \gamma_2 = 0$) and every non-negative function $h$ on $[0,\infty) \times G$ that is parabolic and bounded on $[0, \gamma\phi(2R)] \times B(z, 2R)$,*

$$\sup_{(t,y)\in Q(\delta\phi(R),z,R)} h(t,y) \le c_1 \inf_{y\in B(z,R)} h(0,y).$$

*In particular, the following holds for $t \le 1$ (resp. $t > 0$ when $\gamma_1 = \gamma_2 = 0$).*

$$\sup_{(s,y)\in Q((1-\gamma)t,z,\phi^{-1}(t))} p(s,x,y) \le c \inf_{y\in B(z,\phi^{-1}(t))} p((1+\gamma)t,x,y). \tag{4.32}$$

The proof is basically the same as that of Proposition 4.3 in [11]. See Appendix for details.

### 4.9 Heat kernel lower bound

In this subsection, we prove the lower bound of Theorem 1.2 for $x, y \in F \setminus \mathcal{N}$.

We first derive a near diagonal estimate.

**Lemma 4.13** *There exist $c_1$, $c_2 > 0$ such that*

$$p(t,x,y) \ge c_1 \frac{1}{V(\phi^{-1}(t))}$$

*for all $t \in (0,1]$ (resp. $t > 0$ when $\gamma_1 = \gamma_2 = 0$) and $x, y \in F \setminus \mathcal{N}$ with $\rho(x,y) \le c_2\, \phi^{-1}(t)$.*

**Proof.** In this proof, we will use (1.12) and (1.13) several times without specific mentioning. By Proposition 4.9, there exists $c_3 > 1$ such that for every $t \in (0,1]$ (resp. for $t > 0$ when $\gamma_1 = \gamma_2 = 0$),

$$\mathbb{P}^x \left( \sup_{s\le t} \rho(Y_s, x) > c_3\phi^{-1}(t) \right) \le 1/2.$$

On the other hand, using the upper bound in Theorem 1.2, there exists $0 < c_4 < c_3/2$ such that

$$\mathbb{P}^x(Y_t \in B(x, c_4\phi^{-1}(t))) \le 1/4$$

for every $t \in (0,1]$ (resp. for $t > 0$ when $\gamma_1 = \gamma_2 = 0$). Thus, if we define $E(t) := B(x, c_3\phi^{-1}(t)) \setminus B(x, c_4\phi^{-1}(t))$, then $\mathbb{P}^x(Y_t \in E) \ge 1/4$. Define $t_* = (1-\gamma)t$. Applying the above with $t_*$ in place of $t$ yields $p(t_*, x, z) \ge c_5/V(\phi^{-1}(t_*))$ for some $z \in E(t_*)$, since $\mu(E(t_*)) \le c_6 V(\phi^{-1}(t_*))$. Now by Theorem 4.12, by selecting $c_2 > 0$ sufficiently small, we have that for all $y$ with $\rho(y,x) < c_2\phi^{-1}(t)$,

$$p(t,x,y) \ge \inf_{w\in B(x,c_2\phi^{-1}(t))} p(t,x,w) \ge c_1^{-1} \sup_{(s,w)\in Q(\gamma t,\, x,\, c_2\phi^{-1}(t))} p(t-s,x,w). \tag{4.33}$$

If $c_2 \ge c_3$, the right hand side of (4.33) is greater than or equal to $c_6/V(\phi^{-1}(t))$ and the result holds. If $c_2 < c_3$, by applying Theorem 4.12 iteratively, we can obtain the result similarly. □

*Proof of Theorem 1.2 (Lower bound).* As before, let $t \in (0,1]$ (resp. $t > 0$ when $\gamma_1 = \gamma_2 = 0$). Due to Lemma 4.13, it is enough to prove the theorem for $\rho(x,y) \geq c_2\,\phi^{-1}(t)$. Applying Proposition 4.11 with $t_* = (1-\gamma)t$ in place of $t$, we have

$$\mathbb{P}^x(Y_{t_*} \in B(y,\, c_1\phi^{-1}(t_*))) \geq c_2 \frac{t_* V(\phi^{-1}(t_*))}{V(\rho(x,y))\phi(c_3\rho(x,y))}.$$

As $\mu(B(y,\, c_1\phi^{-1}(t_*))) \leq c_4 V(\phi^{-1}(t_*))$, the above implies $p(t_*, x, z) \geq c_5\, t/(V(\rho(x,y))\phi(c_3\rho(x,y)))$ for some $z \in B(y,\, c_1\phi^{-1}(t_*))$. By applying (4.32) as before, we obtain the desired result. □

### 4.10 Hölder continuity

So far, we have proved Theorem 1.2 for $x, y \in F \setminus \mathcal{N}$. Now we can prove the following Hölder continuity of the heat kernel by applying the estimates obtained above. As a result, we obtain Theorem 1.2 for every $x, y \in F$.

**Proposition 4.14** *For every $R_0 \in (0,1]$ (resp. $R_0 > 0$ when $\gamma_1 = \gamma_2 = 0$), there are constants $c = c(R_0) > 0$ and $\kappa > 0$ such that for every $0 < R \leq R_0$ and every bounded parabolic function $h$ in $Q(0, x_0, 2R)$,*

$$|h(s,x) - h(t,y)| \leq c\,\|h\|_{\infty,F}\, R^{-\kappa}\, \left(\phi^{-1}(|t-s|) + \rho(x,y)\right)^{\kappa} \tag{4.34}$$

*holds for $(s,x)$, $(t,y) \in Q(0,x_0,R)$, where $\|h\|_{\infty,F} := \sup_{(t,y)\in[0,\,\gamma\phi(2R)]\times F} |h(t,y)|$. In particular, for the transition density function $p(t,x,y)$ of $Y$, for any $t_0 \in (0,1)$ (resp. any $T > 0$ and any $t_0 \in (0,T)$ when $\gamma_1 = \gamma_2 = 0$), there are constants $c = c(t_0) > 0$ and $\kappa > 0$ such that for any $t,\, s \in [t_0,\, 1]$ (resp. $t, s \in [t_0, T]$) and $(x_i, y_i) \in F \times F$ with $i = 1, 2$,*

$$|p(s,x_1,y_1) - p(t,x_2,y_2)| \leq c\,\frac{1}{V(\phi^{-1}(t_0))\phi^{-1}(t_0)^{\kappa}}\, \left(\phi^{-1}(|t-s|) + \rho(x_1,x_2) + \rho(y_1,y_2)\right)^{\kappa}. \tag{4.35}$$

The proof is an easy modification of that of Theorem 4.14 in [11], so we will omit it.

## 5 Appendix A: Lévy system

Any Hunt process $X$ on $E$ has a Lévy system $(N, H)$. That is, $N(x,dy)$ is a kernel on $(E_\partial, \mathcal{B}(E_\partial))$ where $E_\partial = E \cup \{\Delta\}$ and $H$ is a PCAF $H$ of $X$ in the strict sense with bounded 1-potential such that for any nonnegative Borel function $f$ on $E \times E_\partial$ that vanishes on the diagonal and is extended to be zero elsewhere,

$$\mathbb{E}_x\left[\sum_{s\leq t} f(X_{s-}, X_s)\right] = \mathbb{E}_x\left[\int_0^t \int_{E_\partial} f(X_s, y) N(X_s, dy)\, dH_s\right] \tag{5.1}$$

for every $x \in E$ and $t \geq 0$.

We claim that $t$ in (5.1) can be replaced by any stopping time $T$ with respect to the minimal filtration $\{\mathcal{F}_t, t \geq 0\}$ generated by $X$.

This is because for any nonnegative bounded Borel function $f$ on $E \times E_\partial$ and $n \geq 1$, $t \mapsto C_t^n := \sum_{s \leq t} f(X_{s-}, X_s) \mathbf{1}_{\{\rho(X_{s-}, X_s) > 1/n\}}$ is a non-decreasing function locally integrable functional and so it admits a dual predictable projection $A_t^n$. That is, for every bounded non-negative predictable process $H_s$,

$$\mathbb{E}_x \left[ \int_0^\infty H_s dC_s^n \right] = \mathbb{E}_x \left[ \int_0^\infty H_s dA_s^n \right].$$

For any stopping time $T$ of $\{\mathcal{F}_t, t \geq 0\}$, taking $H_s = \mathbf{1}_{\{t \leq T\}}(s)$ in the above identity yields

$$\mathbb{E}_x \left[ \sum_{s \leq T} f(X_{s-}, X_s) \mathbf{1}_{\{\rho(X_{s-}, X_s) > 1/n\}} \right] = \mathbb{E}_x \left[ A_T^n \right]. \tag{5.2}$$

Since $X$ is quasi-continuous on $[0, \infty)$, $A^n$ is a PCAF of $X$. According to (5.1) and (5.2),

$$\mathbb{E}_x \left[ \int_0^t \int_{E_\partial} f(X_s, y) \mathbf{1}_{\{\rho(X_s, y) > 1/n\}} N(X_s, dy) dH_s \right] = \mathbb{E}_x \left[ A_t^n \right] \qquad \text{for every } x \in E.$$

In particular, we have for every non-negative excessive function $h$ of $X$,

$$\lim_{t \downarrow 0} \frac{1}{t} \int_E h(x) \mathbb{E}_x \left[ \int_0^t \int_{E_\partial} f(X_s, y) \mathbf{1}_{\{\rho(X_s, y) > 1/n\}} N(X_s, dy) dH_s \right] m(dx) = \lim_{t \downarrow 0} \frac{1}{t} \int_E h(x) \mathbb{E}_x \left[ A_t^n \right] m(dx).$$

It follows from the Revuz identity,

$$\int_E h(x) \int_{E_\partial} f(X_s, y) \mathbf{1}_{\{\rho(X_s, y) > 1/n\}} N(X_s, dy) \mu_H(dx) = \int_E h(x) \mu_{A^n}(dx),$$

where $\mu_{A^n}$ is the Revuz measure of $A^n$. Hence

$$\left( \int_{E_\partial} f(X_s, y) \mathbf{1}_{\{\rho(X_s, y) > 1/n\}} N(X_s, dy) \right) \mu_H(dx) = \mu_{A^n}(dx),$$

which is equivalent to

$$A_t^n = \int_0^t \int_{E_\partial} f(X_s, y) \mathbf{1}_{\{\rho(X_s, y) > 1/n\}} N(X_s, dy) dH_s \qquad \text{for every } t \geq 0.$$

So (5.2) can be rewritten as

$$\mathbb{E}_x \left[ \sum_{s \leq T} f(X_{s-}, X_s) \mathbf{1}_{\{\rho(X_{s-}, X_s) > 1/n\}} \right] = \mathbb{E}_x \left[ \int_0^T \int_{E_\partial} f(X_s, y) \mathbf{1}_{\{\rho(X_s, y) > 1/n\}} N(X_s, dy) dH_s \right]$$

for any stopping time $T$. Passing $n \to \infty$, we conclude that

$$\mathbb{E}_x\left[\sum_{s\le T} f(X_{s-},X_s)\right] = \mathbb{E}_x\left[\int_0^T \int_{E_\partial} f(X_s,y)N(X_s,dy)dH_s\right] \tag{5.3}$$

holds for every bounded $f \ge 0$ that vanishes along the diagonal and for every stopping time $T$. Another truncation argument shows that (5.3) holds for every $f \ge 0$ that vanishes along the diagonal and for every *stopping time* $T$.

Note that in some literature, the Lévy system is simply defined as follows: there is $(N,H)$ such that for every bounded non-negative function that vanishes along the diagonal

$$t \mapsto \int_0^t \int_{E_\partial} f(X_s,y)N(X_s,dy)dH_s$$

is the dual predictable projection of $t \mapsto \sum_{s\le t} f(X_{s-},X_s)$. (In fact, one needs to impose additional condition on $f$ so the local integrability condition $t \mapsto \sum_{s\le t} f(X_{s-},X_s)$ holds. But this can always be circumvented by first working on $f(x,y)\mathbf{1}_{\{\rho(x,y)>\varepsilon\}}$ and then letting $\varepsilon \to 0$, as is done above.) Then (5.3) follows easily from this. But as we see from above that this definition is equivalent to the seemly weaker version (5.1).

# 6 Appendix B: Proof of Parabolic Harnack Inequality

In this appendix, we will prove Theorem 4.12 for $R \le 1$. The extension to $R > 0$ when $\gamma_1 = \gamma_2 = 0$ is straightforward. For the proof, we need the following three lemmas.

**Lemma 6.1** *There exists $C_1 > 0$ such that for every $x \in F$, $r \in (0,1]$, $y \in B(x,r/3)$ and a bounded nonnegative function $h$ on $[0,\infty)\times F$ that is supported in $[0,\infty)\times B(x,2r)^c$,*

$$\mathbb{E}^{(0,x)}\left[h(\tau_r, Y_{\tau_r})\right] \le C_1 \mathbb{E}^{(0,y)}\left[h(\tau_r, Y_{\tau_r})\right], \tag{6.1}$$

*where $\tau_r = \tau_{Q(0,x,r)}$.*

**Proof.** Note that under $\mathbb{P}^{(0,z)}$, $\tau_r := \inf\{t \ge 0 : Y_t \notin B(x,r)\} \wedge \gamma\phi(r) = \tau_{B(x,r)} \wedge \gamma\phi(r)$. Recall that $\gamma := \gamma(1/2,1/2)$. Since $h$ is supported on $(0,\infty)\times B(x,2r)^c$, for any $z \in B(x,r/3)$

$$\mathbb{E}^{(0,z)}\left[h(\tau_r, Y_{\tau_r})\right] = \mathbb{E}^{(0,z)}\left[h(\tau_r, Y_{\tau_r});\ Y_{\tau_r-} \neq Y_{\tau_r}\right].$$

So, by Lemma 4.8 with $T=\tau_r$ and by the fact $r\le 1$,

$$\begin{aligned}
&\mathbb{E}^{(0,z)}\left[h(\tau_r, Y_{\tau_r})\right] \\
= \;& \mathbb{E}^{(0,z)}\left[\int_0^{\tau_r}\int_{B(x,2r)^c} h(s,u)J(Y_s,u)\,\mu(du)\,ds\right] \\
\asymp \;& \mathbb{E}^{(0,z)}\left[\int_0^{\tau_r}\int_{B(x,2r)^c} \frac{h(s,u)}{V(\rho(x,u))\phi(\rho(x,u))}\,\mu(du)\,ds\right] \\
= \;& \int_0^{\gamma\phi(r)}\int_{B(x,2r)^c} \frac{h(s,u)\,\mathbb{P}^{(0,z)}(s<\tau_r)}{V(\rho(x,u))\phi(\rho(x,u))}\,\mu(du)\,ds.
\end{aligned}$$

On the other hand, by Proposition 4.9, for all $s\le\gamma\phi(r)$ and $z\in B(x,r/3)$,

$$1/2\le \mathbb{P}^z(s\le\tau_{B(z,r/2)})\le \mathbb{P}^{(0,z)}(s\le\tau_r)\le 1.$$

This implies that the values of the function $z\mapsto\mathbb{E}^z\left[h(\tau_r,Y_{\tau_r})\right]$ are all comparable with each other with a universal constant multiple for any $z\in B(x,r)$, and therefore proves the lemma. $\square$

For each $A\subset[0,\infty)\times F$, denote $\sigma_A:=\inf\{t>0: Z_t\in A\}$ and $A_s:=\{y\in F:\ (s,y)\in A\}$.

**Lemma 6.2** *There exists $C_2>0$ such that for all $x\in F$, $r\in(0,1]$ and any compact subset $A\subset Q(0,x,r)$,*

$$\mathbb{P}^{(0,x)}(\sigma_A<\tau_r)\ge C_2\frac{m\otimes\mu(A)}{V(r)\phi(r)},$$

*where $\tau_r=\tau_{Q(0,x,r)}$ and $m\otimes\mu$ is a product measure of the Lebesgue measure $m$ on $\mathbb{R}_+$ and $\mu$ on $F$.*

**Proof.** The conclusion of the lemma clearly holds if $\mathbb{P}^x(\sigma_A<\tau_r)>1/4$. So we will assume $\mathbb{P}^x(\sigma_A<\tau_r)\le 1/4$. Let $T=\sigma_A\wedge\tau_r$. Then

$$\mathbb{P}^{(0,x)}(\sigma_A<\tau_r)=\mathbb{P}^{(0,x)}((T,Y_T)\in A)\ge\mathbb{P}^{(0,x)}((T,Y_T)\in A;\ Y_{T-}\ne Y_T).$$

Applying Lemma 4.8 with $f(s,x,y)=\mathbf{1}_{\{x\ne y\}}\mathbf{1}_A(s,y)$ and $T=\sigma_A\wedge\tau_r$, we have, with $t_0:=\gamma\phi(r)$ and $r\in(0,1]$,

$$\begin{aligned}
\mathbb{P}^{(0,x)}(\sigma_A<\tau_r) \;&\ge\; \mathbb{E}^{(0,x)}\left[\int_0^T\left(\int_{A_s}\frac{c_1}{V(\rho(u,Y_s))\phi(c_2\rho(u,Y_s))}\mu(du)\right)ds\right] \\
&\ge\; c_1\mathbb{E}^{(0,x)}\left[\int_0^{t_0}\int_{A_s}\frac{1}{V(2r)\phi(2c_2r)}\,\mu(du)\,ds;\ \sigma_A\wedge\tau_r\ge t_0\right] \\
&\ge\; c_1\frac{m\otimes\mu(A)}{V(r)\phi(r)}\,\mathbb{P}^{(0,x)}(\sigma_A\wedge\tau_r\ge t_0).
\end{aligned}$$

Here in the first inequality, we used the fact that $\rho(u,y) \leq 2r$ for $y \in B(x,r)$ and $u \in A_s \subset B(x,r)$, and we used (1.12)-(1.13) in the last inequality. By Proposition 4.9,

$$\mathbb{P}^{(0,x)}(\tau_r < t_0) \leq \mathbb{P}^x(\tau_{B(x,r/2)} < \gamma\phi(r)) \leq 1/2,$$

and so

$$\mathbb{P}^{(0,x)}(\sigma_A \wedge \tau_r \geq t_0) \geq 1 - \mathbb{P}^{(0,x)}(\sigma_A < \tau_r) - \mathbb{P}^{(0,x)}(\tau_r < t_0) \geq 1/4,$$

which proves the lemma. □

Define $U(t,x,r) = \{t\} \times B(x,r)$.

**Corollary 6.3** *For every $0 < \delta \leq \gamma$, there exists $C_3 > 0$ such that for every $R \in (0,1]$, $r \in (0, R/4]$ and $(t,x) \in Q(0,z,R/3)$ with $t \geq \delta\phi(r)$,*

$$\mathbb{P}^{(0,z)}\left(\sigma_{U(t,x,r)} < \tau_{Q(0,z,R)}\right) \geq C_3 \frac{V(r)\phi(r)}{V(R)\phi(R)}.$$

**Proof.** Let $Q' := [t - \delta\phi(r),\, t] \times B(x, r/2) \subset Q(0,z,R/2)$. By Lemma 6.2,

$$\mathbb{P}^{(0,z)}\left(\sigma_{Q'} < \tau_{Q(0,z,R)}\right) \geq c_1 \delta V(r)\phi(r)/(V(R)\phi(R)).$$

Starting from any point in $Q'$, by Proposition 4.9 there is a probability at least $1/2$ that the process $Y$ stays in $B(x,r/2)$ for at least $\gamma\phi(r)$ amount of time. Thus, by the strong Markov property, with probability at least $\delta\, V(r)\phi(r)/(2V(R)\phi(R))$, the process hits $Q'$ before exiting $Q(0,z,R)$ and stays within $B(x,r)$ for an additional $\gamma\phi(r)$ amount of time, and hence hits $U(t,x,r)$ before exiting $Q(0,z,R)$. □

Recall that $Z_s = (V_s, Y_s)$ is the space-time process of $Y$, where $V_s = V_0 + s$. The following is a standard fact proved in Lemma 4.13 in [11].

**Lemma 6.4** (Lemma 4.13 in [11]) *For any bounded Borel measurable function $q(t,x)$ that is parabolic in an open subset $D$ of $\mathbb{R}_+ \times F$, $s \mapsto q(Z_{s\wedge\tau_D})$ is right continuous $\mathbb{P}^{(t,x)}$-a.s. for every $(t,x) \in D$. Here $\tau_D = \inf\{s > 0 : Z_s \notin D\}$.*

**Proof of Proposition 4.12.** Let $R \in (0,1]$. By Lemma 6.1 and its proof, we see that $\inf_{y\in B(z,R)} h(0,y) > 0$ unless $h$ is identically zero. Taking a constant multiple of $h$ if needed, we may assume that

$$\inf_{y\in B(z,R)} h(0,y) = 1/2.$$

Let $v \in B(z,R)$ be such that $h(0,v) \le 1$. It is enough to show that $h(t,x)$ is bounded from above in $Q(\delta\phi(R), z, R)$ by a constant that is independent of the function $h$.

By Lemma 6.2, there exists $c_1 \in (0,1/2)$ such that if $r \le R/2$ and $D \subset Q(t,x,r)$ having $m \otimes \mu(D)/m \otimes \mu(Q(t,x,r)) \ge 2/3$, then

$$\mathbb{P}^{(t,x)}(\sigma_D < \tau_r) \ge c_1, \tag{6.2}$$

where $\tau_r := \tau_{Q(t,x,r)}$. Define

$$\eta = \frac{c_1}{3} \quad \text{and} \quad \xi = \frac{1}{3} \wedge (C_1^{-1}\eta), \tag{6.3}$$

where $C_1$ is the constant in Lemma 6.1. We claim that there is a universal constant $K = K(\delta)$ to be determined later, which is independent of $R$ and function $h$, such that $h \le K$ on $Q(\delta\phi(R), z, R)$. We are going to prove this by contradiction. Suppose this is not true. Then there is some point $(t_0, x_0) \in Q(\delta\phi(R), z, R)$ such that $h(t_0, x_0) > K$. We will show that there is a constant $\beta > 0$ and there is a sequence of points $\{(t_k, x_k)\}$ in $\widehat{Q}(0, z, R) := [0, \gamma\phi(2R)] \times B(z, 2R)$ so that $h(t_k, x_k) \ge (1+\beta)^k K$, which contradicts to the assumption that $h$ is bounded on $\widehat{Q}(0, z, R)$.

Let $r > 0$ to be the smallest $r$ such that

$$\frac{m \otimes \mu(Q(0, x_0, r))}{V(R)\phi(R)} \ge \frac{3}{C_2\, \xi K}, \qquad \text{and} \qquad \frac{V(r)\phi(r)}{V(R)\phi(R)} \ge \frac{2}{C_3\, \xi K}, \tag{6.4}$$

where $C_2$ and $C_3$ are the constants in Lemma 6.2 and Corollary 6.3 respectively. With $K$ being sufficiently large, such $r$ exists and can be made to be less than $R/4$. This is because by using (1.11)-(1.12) and the fact that $0 < r < R \le 1$ in the case of $\gamma_2 \ge \gamma_1 > 0$, we see that there exists $C > 0$ such that

$$\frac{C}{K} \ge \frac{V(r)\phi(r)}{V(R)\phi(R)} \ge \left(\frac{r}{R}\right)^{d+\beta_2}. \tag{6.5}$$

Thus $r \le C' R K^{-1/(d+\beta_2)}$. So we can take $K > 1$ large enough such that $C' K^{-1/(d+\beta_2)} < 1/4$ and let

$$r := C' R K^{-1/(d+\beta_2)}. \tag{6.6}$$

Let $U = \{t_0\} \times B(x_0, r)$. Were function $h \ge \xi K$ on $U$, we would have by Corollary 6.3 that for $v \in B(x_0, r)$,

$$1 \ge h(0,v) = \mathbb{E}^{(0,v)}\left[h(Z_{\sigma_U \wedge \tau_Q})\right] \ge \xi K\, \mathbb{P}^{(0,v)}\left(\sigma_U < \tau_Q\right) \ge \xi K \, \frac{C_3 V(r)\phi(r)}{V(R)\phi(R)},$$

where $Q := Q(0, x_0, R)$, which contradicts to our choice of $r$ in (6.4). Thus,

$$\text{there must be at least one point in } U \text{ on which } h \text{ takes a value less than } \xi K. \tag{6.7}$$

We next claim that

$$\mathbb{E}^{(t_0,x_0)}\left[h(\tau_r, Y_{\tau_r}) : Y_{\tau_r} \notin B(x_0, 2r)\right] \leq \eta K, \tag{6.8}$$

where $\tau_r := \tau_{Q(t_0,x_0,r)}$. If not, then by Lemma 6.1, we would have for every $y \in B(x_0, r)$,

$$\begin{aligned} h(t_0, y) &\geq \mathbb{E}^{(t_0,y)}[h(\tau_r, Y_{\tau_r}) : Y_{\tau_r} \notin B(x_0, 2r)] \\ &\geq C_1^{-1}\mathbb{E}^{(t_0,x_0)}[h(\tau_r, Y_{\tau_r}) : Y_{\tau_r} \notin B(x_0, 2r)] \\ &\geq C_1^{-1}\eta K \geq \xi K, \end{aligned}$$

a contradiction to (6.7). So (6.8) holds.

Let $A$ be any compact subset of

$$\widetilde{A} := \{(s, y) \in Q(t_0, x_0, r) : h(s, y) \geq \xi K\}.$$

By Lemma 6.2

$$1 \geq h(0, x_0) \geq \mathbb{E}^{(0,x_0)}[h(Z_{\sigma_A}) : \sigma_A < \tau_Q] \geq \xi K \mathbb{P}^{(0,x_0)}(\sigma_A < \tau_Q) \geq \xi K \, \frac{C_2 \, m \otimes \mu(A)}{V(R)\phi(R)}.$$

So by (6.4),

$$\frac{m \otimes \mu(A)}{m \otimes \mu(Q(t_0, x_0, r))} \leq \frac{V(R)\phi(R)}{C_2 \xi K \cdot m \otimes \mu(Q(t_0, x_0, r))} \leq \frac{1}{3}. \tag{6.9}$$

Since (6.9) holds for every compact subset $A$ of $\widetilde{A}$, it holds for $\widetilde{A}$ in place of $A$.

Let $D = Q(t_0, x_0, r) \setminus \widetilde{A}$ and $M = \sup_{(s,y)\in Q(t_0,x_0,2r)} h(s, y)$. We write

$$\begin{aligned} h(t_0, x_0) &= \mathbb{E}^{(t_0,x_0)}[h(\sigma_D, Y_{\sigma_D}) : \sigma_D < \tau_r] + \mathbb{E}^{(t_0,x_0)}[h(\tau_r, Y_{\tau_r}) : \tau_r \leq \sigma_D, Y_{\tau_r} \notin B(x_0, 2r)] \\ &\quad + \mathbb{E}^{(t_0,x_0)}[h(\tau_r, Y_{\tau_r}) : \tau_r \leq \sigma_D, Y_{\tau_r} \in B(x_0, 2r)]. \end{aligned}$$

The first term on the right is bounded by $\xi K \, \mathbb{P}^{(t_0,x_0)}(\sigma_D < \tau_r)$ in view of Lemma 6.4, the second term is bounded by $\eta K$ according to (6.8), and the third term is clearly bounded by $M\mathbb{P}^{(t_0,x_0)}(\tau_r \leq \sigma_D)$. Recall that $h(t_0, x_0) > K$. Therefore,

$$K \leq \xi K \, \mathbb{P}^{(t_0,x_0)}(\sigma_D < \tau_r) + \eta K + M \, \mathbb{P}^{(t_0,x_0)}(\sigma_D \geq \tau_r).$$

It follows from (6.9) and (6.2)-(6.3)

$$M/K \geq \frac{1 - \eta - \xi \, \mathbb{P}^{(t_0,x_0)}(\sigma_D < \tau_r)}{\mathbb{P}^{(t_0,x_0)}(\sigma_D \geq \tau_r)} \geq \frac{1 - \eta - \xi}{1 - c_1} + \xi = \frac{1 - \eta - \xi \, c_1}{1 - c_1} \geq \frac{1 - (2c_1)/3}{1 - c_1} := 1 + \beta,$$

where $\beta = \frac{c_1}{3(1-c_1)}$. In other words, $M \geq (1 + \beta)K$. As $M = \sup_{(s,y)\in Q(t_0,x_0,2r)} h(s, y)$, there exists a point $(t_1, x_1) \in Q(t_0, x_0, 2r) \subset \widehat{Q}(0, z, R)$ such that $h(t_1, x_1) \geq (1 + \beta)K =: K_1$.

We now iterate the above procedure to obtain a sequence of points $\{(t_k, x_k)\}$ in the following way. Using the above argument (with $(t_1, x_1)$ and $K_1$ in place of $(t_0, x_0)$ and $K$), there exists $(t_2, x_2) \in Q(t_1, x_1, 2r_1)$ such that

$$r_1 = C'RK_1^{-1/(d+\beta_2)} = C'(1+\beta)^{-1/(d+\beta_2)}K^{-1/(d+\beta_2)}\,R$$

and $h(t_2, x_2) \geq (1+\beta)K_1 = (1+\beta)^2 K =: K_2$. We continue this procedure to obtain a sequence of points $\{(t_k, x_k)\}$ such that with

$$r_k := C'RK_k^{-1/(d+\beta_2)} = C'(1+\beta)^{-k/(d+\beta_2)}K^{-1/(d+\beta_2)}\,R,$$

$(t_{k+1}, x_{k+1}) \in Q(t_k, x_k, 2r_k)$ and $h(t_{k+1}, x_{k+1}) \geq (1+\beta)^{k+1}K =: K_{k+1}$. As $0 \leq t_{k+1} - t_k \leq \gamma\phi(2r_k)$ and $\rho(x_{k+1}, x_k) \leq 2r_k$, by (6.5), we can take $K$ large enough (independent of $R$ and $h$) so that $(t_k, x_k) \in \widehat{Q}(0, z, R)$ for all $k$. This is a contradiction because $h(t_k, x_k) \geq (1+\beta)^k K$ goes to infinity as $k \to \infty$ while $h$ is bounded on $\widehat{Q}(0, z, R)$. We conclude that $h$ is bounded by $K$ in $Q(\delta\phi(R), z, R)$, which completes the proof of the proposition. $\square$

**Acknowledgement.** The authors are grateful to Martin Barlow and Alexander Grigor'yan for valuable comments and suggestions.

## References

[1] M. T. Barlow, R. F. Bass, Z.-Q. Chen and M. Kassmann, Non-local Dirichlet forms and symmetric jump processes. Preprint 2006.

[2] M. T. Barlow, A. Grigor'yan and T. Kumagai, Heat kernel upper bounds for jump processes. In preparation.

[3] R. F. Bass and M. Kassmann, Harnack inequalities for non operators of variable order. *Trans. Amer. Math. Soc.,* **357** (2005), 837–850.

[4] R. F. Bass and M. Kassmann. Hölder continuity of harmonic functions with respect to operators of variable orders. *Comm. Part. Diff. Eq.,* **30** (2005), 1249–1259.

[5] R. F. Bass and D. A. Levin, Harnack inequalities for jump processes. *Potential Anal.,* **17** (2002), 375–388.

[6] R. F. Bass and D. A. Levin, Transition probabilities for symmetric jump processes. *Trans. Amer. Math. Soc.,* **354** (2002), 2933–2953.

[7] K. Bogdan, A. Stós, and P. Sztonyk, Potential theory for Lévy stable processes. *Bull. Polish Acad. Sci. Math.,* **50** (2002), 361–372.

[8] K. Bogdan and P. Sztonyk. Harnack's inequality for stable Lévy processes. *Potential Anal.,* **22** (2005), 133–150.

[9] E. A. Carlen and S. Kusuoka and D. W. Stroock, Upper bounds for symmetric Markov transition functions. *Ann. Inst. Heri. Poincaré-Probab. Statist.,* **23** (1987), 245-287.

[10] Z.-Q. Chen, On reflected Dirichlet spaces, *Probab. Theory Relat. Fields,* **94** (1992), 135-162.

[11] Z.-Q. Chen and T. Kumagai, Heat kernel estimates for stable-like processes on $d$-sets. *Stochastic Process Appl.,* **108** (2003), 27-62.

[12] Z.-Q. Chen and R. Song, Drift transforms and Green function estimates for discontinuous processes. *J. Funct. Anal.,* **201** (2003), 262-281.

[13] T. Coulhon, Ultracontractivity and Nash type inequalities. *J. Funct. Anal.,* **141** (1996), 510-539.

[14] M. Fukushima, Y. Oshima and M. Takeda, *Dirichlet forms and symmetric Markov processes.* de Gruyter, Berlin, 1994.

[15] I. W. Herbst and D. Sloan, Perturbation of translation invariant positive preserving semigroups on $L^2(\mathbb{R}^N)$. *Trans. Amer. Math. Soc.,* **236** (1978), 325-360.

[16] J. Hu and T. Kumagai, Nash-type inequalities and heat kernels for non-local Dirichlet forms. *Kyushu J. Math.,* **60** (2006), 245–265.

[17] N. Ikeda, N. Nagasawa, and S. Watanabe, A construction of Markov processes by piecing out. *Proc. Japan Acad.,* **42** (1966), 370–375.

[18] V. Kolokoltsov, Symmetric stable laws and stable-like jump-diffusions. *Proc. London Math. Soc.* **80** (2000), 725–768.

[19] K. Kuwae and M. Takahashi, Kato class functions of Markov processes under ultracontractivity. To appear *Adv. Studies Pure Math.*, Math. Soc. Japan.

[20] P.-A. Meyer. Renaissance, recollements, mélanges, ralentissement de processus de Markov. *Ann. Inst. Fourier*, **25** (1975), 464–497.

[21] R. Schilling and T. Uemura, Dirichlet forms generated by pseudo differential operators: on the Feller property of the associated stochastic process. Preprint 2005.

[22] R. Song and Z. Vondraček, Harnack inequality for some classes of Markov processes. *Math. Z.*, **246** (2004), 177–202.

Zhen-Qing Chen
Department of Mathematics, University of Washington, Seattle, WA 98195, USA
E-mail: zchen@math.washington.edu

Takashi Kumagai
Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606-8502, Japan
E-mail: kumagai@kurims.kyoto-u.ac.jp